Role of STAG2 in Myeloid Malignancies

Matto, Nazia

Awarding institution:
King's College London

Role of *STAG2* in Myeloid Malignancies

University of London

Nazia Matto

THESIS PRESENTED FOR THE DEGREE

OF

DOCTOR OF PHILOSPHY

King's College London

Department of Haematological Medicine

2017

Declaration

I hereby declare that I alone composed this thesis and that the work presented here is my own, except where stated otherwise.

Nazia Matto
2017

Acknowledgement

I would like to express my deepest gratitude to Prof Ghulam Jeelani Mufti for his guidance, encouragement and endless support during my PhD. Thank you for having the faith in me and giving me this opportunity. It has been an honour to be your student. I would like to acknowledge the valuable input of my second supervisor Dr Terry Gaymes who through his skilful guidance, innovative ideas and patience contributed to many discussions that help shape this project.

Many thanks to our ET group members Joop, Alex, Azim, Jiang and Andrea for their co-operation and support and also to members of tissue bank Rajani and Nigel for helping me sorting out the patient samples. I am very much thankful to Dr David Darling for helping me generating the lentivirus. To Syed for all his help during these years. This PhD gave rise to an ever lasting bond of friendship within our small girls group Sneha, Heba and Shok ping with whom I have shared endless moments of joy and sorrow. Thank you for being there for me.

I would like to express my appreciation to my PhD co-ordinators Prof Shaun Thomas and Dr Alan Ramsay for their guidance and feedback. I am grateful to Dr Austin Kulasekararaj for providing me with patient clinical details despite his busy schedule.

Last but not the least my pillar of support my family, my parents and siblings, Mahreen, Tazeen and Abid for their endless prayers, support and encouragement all through these years. A huge thankyou to Moyeen, Rabia and my little nephew Dawood for their love and support during this journey.

Abstract

Myelodysplastic syndrome (MDS) a clonal disease arising from the mutation of the haematopoietic stem cells is characterized by the presence of peripheral blood cytopenias, dysplastic haematopoietic differentiation and transformation to acute myeloid leukaemia (AML). Defects in the DNA damage repair pathway in MDS and AML give rise to chromosomal instability (CI) such as aneuploidy, chromosomal translocation and loss of heterozygosity (LOH) a hallmark of human neoplasm. Cohesin a multiprotein complex responsible for tethering the sister chromatids together until their timely separation in mitosis has been linked with the phenomena of chromosomal instability. In this study, second generation sequencing (SGS) was performed in a panel of 154 MDS patients to characterize the incidence and functional consequences of cohesin mutations in myeloid malignancies. This led to the identification of mutations in three components of cohesin ring in 6% (9/154) of the patients with maximum number of mutations identified in *STAG2*. None of the patients with cohesin aberrations had a *TP53* mutation and all the patients with cohesin mutations showed a significant progression free survival. Four of the patients showed co-existing mutations in *ASXL1* (an epigenetic regulator) and *SRSF2* (pre-mRNA splicing factor).

To elucidate the functional consequences of *STAG2* mutations and the possibility for it to be exploited synthetically as a target for therapeutic intervention two cell lines; a myeloid cell line, U937 with a stable knockdown of *STAG2* and a urinary bladder cell line, UMUC3 with mutated *STAG2* were used to study the effect of loss of *STAG2* on homologous recombination (HR) repair of DNA, cell cycle kinetics and viability in response to DNA repair pathway inhibitors such as inhibitors of poly-ADP-ribose (PARP) and effect on the expression of other members of the cohesin

complex. Knockdown of *STAG2* had no effect on the recruitment of RAD51 a marker for HR, cell cycle kinetics and viability in response to treatment with PARP inhibitors. Further no effect on the expression of other members of the cohesin complex was observed upon knockdown of *STAG2*. The results from this study speculate *STAG2* as a candidate which has a role to play in the pathogenesis of MDS not through its involvement in HR pathway but through a mechanism that needs further investigation.

Table of Contents

Abbreviations

ASXL1	Additional sex coombs like 1
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3 related
BRCA	Breast Cancer
CdLS	Cornelia de Lange Syndrome
ChIp	Chromatin immunoprecipitation
CML	Chronic myeloid leukaemia
CMML	Chronic myelomonocytic leukaemia
CTCF	CCCTC binding factor
DMSO	Dimethyl sulfoxide
DNA	Deoxyribo nucleic acid
FACS	Fluorescence activated cell sorting
EZH2	Enhancer of zeste
HR	Homologous Recombination
IPSS	International prognostic scoring system
MDS	Myelodysplastic Syndrome
MRN	MRE-RAD50-NBS1
NADP	Nicotinamide adenine dinucleotide phosphate
NHEJ	Non homologous end joining
PARP	Poly-ADP-ribose polymerase
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDSB	Programmed double strand break

PARPi	PARP Inhibitor
RA	Refractory Anaemia
RARS	Refractory Anaemia with excess blast
RARS	Refractory Anaemia with ringed sideroblast
RCMD	Refractory cytopenia with multilineage dysplasia
RNA	Ribose nucleic acid
RPMI	Roswell Park Memorial Institute Medium
sAML	secondary AML
SA	Stromal Antigen
SCC	Sister chromatid cohesion
SGS	Second generation sequencing
SRSF2	serine-arginine rich splicing factor3
T.E	Tris-EDTA
WHO	World Health Organisation
WGS	Whole genome sequencing
XRCC1	X-ray cross complementation 1

List of Figures

List of Tables

Chapter One

INTRODUCTION

1.1 Myelodysplastic Syndrome

Myelodysplastic syndromes (MDSs) are a group of clonal haematopoietic stem cell disorders characterized by ineffective haemopoiesis and transformation to acute myeloid leukaemia (AML). The sequel of MDS results from underlying cytopenias and includes thrombocytopenia, neutropenia and anaemia. The incidence of MDS increases with age and the median age at diagnosis is 70-75 years (Scott and Deeg, 2010). There is increasing evidence that haploinsufficiency, mutations of splicesome and epigenetic genes as well as abnormalities in cytokines, the immune system and bone marrow stroma all contribute to the development of the full spectrum of Myelodysplastic syndromes.

1.2 Classification

Over the years, several MDS classification and prognostic scoring system have been developed to predict the survival or transition to AML, as the prognosis of MDS is highly variable. (Bennett et al., 1982, Mufti et al., 2008, Tefferi and Vardiman, 2009). The French-American-British (FAB) diagnostic schema has been largely supplemented by the World Health Organization (WHO) 2001 criteria (Table 1.1). The revision included lowering the threshold for the percentage of blasts required to make the diagnosis of AML from 30% to 20% thus elimination of the MDS subcategory of refractory anaemia with excess blasts in transformation (RAEB-T). In addition chronic myelomonocytic leukaemia (CMML) was reclassified from a subcategory of MDS to a subcategory of myelodysplastic /myeloproliferative disorder. A revised updated version of WHO classification of MDS came out in 2016 with a small number of changes. In low risk MDS RCMD-RS RS has been removed from RCMD to recognize the importance of *SF3B1* mutations. The general term myelodysplastic syndrome has been associated with single or multilineage dysplasia as the type of dysplasia and cytopenia are not

always concordant. To emphasize the importance of the percentage of blasts that dictate therapy in high risk patients the term Refractory anaemia with excess blasts 1 and 2 have been replaced by myelodysplastic syndrome excess blast 1 and 2.

Table 1.1: WHO diagnostic classification of Myelodysplastic Syndrome

Disease	Blood findings	Bone marrow findings
Refractory anaemia (RA)	**Anaemia No or rare blasts**	**Erythroid dysplasia only <5% blasts <15% ringed sideroblast**
Refractory anaemia with ringed sideroblasts (RARS)	**Anaemia No blasts**	**Erythroid dysplasia only ≥ 15% ringed sideroblasts <5% blasts**
Refractory cytopenia with multilineage dysplasia (RCMD)	**Cytopenia (bicytopenia or pancytopenia) No or rare blasts No Auer rods <1×10^9/L of monocytes**	**Dysplasia in ≥10% of cells in 2 or more myeloid cell lines. <5% of blasts in the marrow No Auer rods <15% ringed sideroblasts**
Refractory cytopenia with multilineage dysplasia and ringed sideroblasts (RCMD-RS)	**Cytopenia (bicytopenia or pancytopenia) No or rare blasts No Auer rods <1×10^9/L of monocyte**	**Dysplasia in ≥10% of cells in two or more myeloid cell lines ≥15% ringed sideroblasts <5% blasts No Auer rods**
Refractory anaemia with excess blasts-1(RAEB-1)	**Cytopenias <5% blasts No Auer rods <1×10^9/L of monocytes**	**Unilineage or multilineage dysplasia 5% to 9% blasts No Auer rods**
Refractory anaemia with excess blasts-2	**Cytopenias 5% to 19% blasts Auer rods ± <1×10^9/L of monocytes**	**Unilineage or multilineage dysplasia 10% to 19% blasts Auer rods ±**
Myelodysplastic syndrome, unclassified (MDS-U)	**Cytopenias No or rare blasts No Auer rods**	**Unilineage dysplasia in granulocytes or megakaryocytes <5% blasts No Auer rods**

The International Prognostic Scoring System (IPSS) was developed by *Greenberg et al* 1997 to provide an improved method for evaluating prognosis in MDS. Variables like cytogenetic abnormalities, percentage of myeloblast in bone marrow

and number of cytopenia offered prognostic information and divided patients into four risk categories Low, Intermediate-1 (INT-1), Intermediate-2 (INT-2), and High (Greenberg et al., 1997) (Table 1.2).

Table 1.2: IPSS Prognostic Scoring System (IPSS) classification criteria. (a) Good, normal, -Y, del (5q), del(20q); Poor, complex (≥ 3 abnormalities) or chromosome 7 anomalies; Intermediate, other abnormalities. (b) Age related survival and progression to AML within the IPSS subgroup.

(a)

Score					
Prognostic variable	0	0.5	1	1.5	2
Bone marrow blasts %	<5	5-10		11-20	21-30
Karyotype	Good	Intermediate	Poor		
Cytopenias	0 to1	2 to3			

(b)

Score	Risk group	Median survival (yr)	AML evolution (yr)
0	Low	5.7	9.4
0.5-1	INT-1	3.5	3.3
1.5-2	INT-2	1.2	1.1
≥ 2.5	High	0.4	0.2

Limitations of IPSS was the exclusion of patients with secondary MDS and inability to estimate the real time risk as only the variables present at the time of diagnosis were included. The IPSS was revised in 2012 to IPSSR. The novel components included 5 rather than 3 cytogenetic prognostic subgroups, splitting the lower marrow blast percentage value and depth of cytopenias. As compared to

IPSS, this model defined 5 major prognostic categories rather than four: Very low, Low, Intermediate, High, Very high (Table 1.3).

Table 1.3: IPSS Prognostic Scoring System Revised (IPSSR) classification criteria. (a) (Very good, -Y, del(11q); Good, Normal, del(5q), del(12p), del(20q), double including del(5q); Intermediate del(7q), +8, +19, i(17q), any other single or double independent clones; Poor -7 inv(3)/t(3q)/del(3q), double including -7/del(7q) complex: 3 abnormalities; Very poor Complex: >3 abnormalities. (b) Age related survival and progression to AML within the IPSS-R subgroup.

(a)

Prognostic variable	0	0.5	1	1.5	2	3	4
Cytogenetics	Very good	-	Good	-	Intermediate	Poor	Very poor
BM blast%	≤ 2%	-	>2%-<5%	-	5%-10%	≥10%	-
Haemoglobin g/dL	≥ 10	-	8-< 10	<8	-	-	-
Platelets	≥100	50-< 100	< 50	-	-	-	-
ANC (Absolute neutrophil count) × 10^9/L	≥ 0.8	< 0.8	-	-	-	-	-

(b)

Score	Risk group	Median survival (yr)	AML evolution (yr)
≤ 1.5	Very Low	8.8	NR
> 1.5-3	Low	5.3	10.8
> 3-4.5	Intermediate	3.0	3.2
> 4.5-6	High	1.6	1.4
> 6	Very High	0.8	0.73

1.2.1 Cytogenetic Abnormalities

Cytogenetic abnormalities are the major determinants in the pathogenesis, diagnosis, and prognosis in MDS and are increasingly becoming the basis of drug selection in the treatment of MDS. Chromosomal abnormalities are seen in 50% of

patients with primary MDS and 80% of cases with therapy related MDS. Cytogenetics plays an important role on the outcome of patients with MDS and AML undergoing haematopoietic stem cell transplant. In AML and chronic myelogenous leukaemia (CML) balanced cytogenetic abnormalities like reciprocal translocation, inversion and insertion are common while unbalanced chromosomal abnormalities such as gain or loss of chromosome are common in MDS. Cytogenetic abnormalities are important because chromosomal regions contain genes that are important in MDS biology and abnormalities such as chromosomal deletions show that haploinsufficiency or loss of tumour suppressor genes (TSG) are important in MDS pathogenesis.

1.2.2 Chromosome 5q deletion

In patients with MDS, deletion of 5q (5q-) is the most common cytogenetic abnormality with an incidence of 15% with isolated 5q- being usually associated with a favourable prognosis. 5q- deletions are universally heterozygous. Two commonly deleted regions (CDRs) 5q33.1 and 5q31 have been found in patients with deletion of 5q-. 5q33.1 is the most distal CDR associated with a phenotype called 5q- syndrome and is characterized with macrocytic anaemia, relative thrombocythemia, female predominance and good prognosis while the proximal CDR at 5q31 is associated with therapy-related MDS and is associated with poor prognosis (Van den Berghe et al., 1974). Many genes are localised to the deleted regions on chromosome 5q. In the last decade significant progress has been made to unravel the molecular mechanism of 5q-syndrome. Using a RNA-mediated interference (RNAi) based approach in $CD34^{+}$ cells *Ebert et al* found haploinsufficiency of ribosomal protein S14 (*RPS 14*) a component of 40S

ribosomal subunit located on 5q32/33 CDR to be associated with the development of anaemia that characterizes the 5q-syndrome (Ebert et al., 2008). However haploinsufficiency of *RPS 14* alone does not recapitulate all the features of 5q-syndrome and thrombocytosis observed in some patients with 5q-syndrome has been linked to deficiency of two microRNA genes miR-145 and miR-146a that map within and adjacent to CDR (Kumar, 2009, Starczynowski et al., 2010). Narrowing down of the CDR in the 5q- syndrome to 1.5 Mb at 5q32 was made by *Boultwood et al*. In this study reduced gene expression approximating to haploinsufficiency was found in CD34+ cells of several candidate genes and included *RPS 14,* secreted protein acidic and rich in cysteine (*SPARC*) a tumour suppressor protein and casein kinase 1 alpha 1 (*CSNK1A1*) a family of serine threonine kinases which play an important role in regulation of hedgehog signalling by governing cell growth and patterning in animal development (Boultwood et al., 2002). Further *Lia et al* identified 1-1.5 Mb CDR at 5q31 in AML and advanced MDS, which is different from the CDR of 5q- syndrome. The CDR at 5q31 encompasses several genes like Early growth response gene (*EGR1*) a transcriptional factor and Alpha catenin (*CTNNA1*) a putative tumor suppressive gene (TSG) (Lai et al., 2001, Joslin et al., 2007). Deletions of 5q are not restricted to CDR but encompass both of these regions and beyond. *NPM1* (Nucleophosmin 1) and *APC* (Anaphase promoting complex) are other genes found outside the CDR and are often lost with deletions of 5q (Grisendi et al., 2005, Wang et al., 2010). Deletion of some genes may not be involved in MDS but could sensitize cells to therapeutic agents like lenalidomide. Two cell cycle regulating phosphatases like cell division cycle 25C (*CDC25C*) and protein phosphatase 2 (*PP2A*) have been implicated in the favourable response to

lenalidomide (Wei et al., 2009). 5q- syndrome is also considered as a disease of haematopoietic stem cell (Gordillo et al., 2008) (Fig1.1).

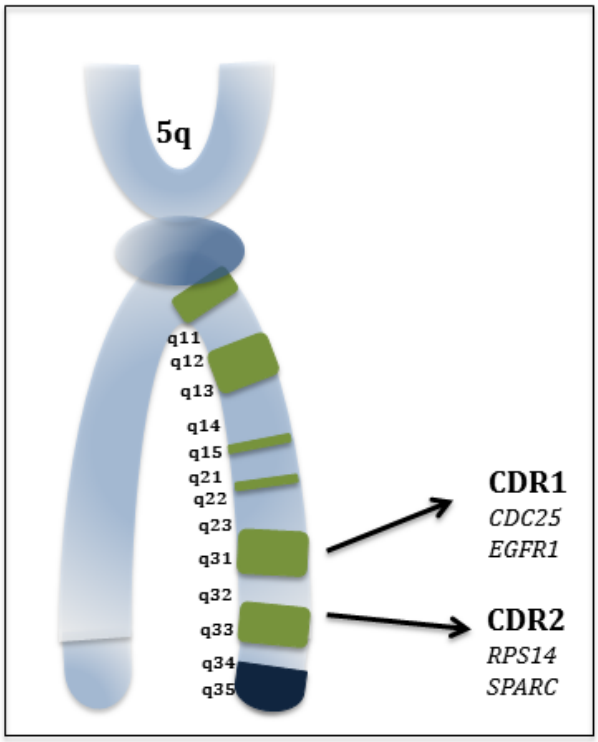


Figure 1.1: Mapping of CDR region on chromosome 5. Location of commonly deleted regions (CDR) on long arm of chromosome 5 in humans. Major genes like *RPS14* a component of 40s ribosomal subunit which is associated with 5q syndrome is located on distal arm at position 5q32/33 while the proximal arm contain genes like early growth response factor 1 a transcriptional factor (*EGRFR1*) and a phosphatase cell division cycle 25 (*CDC25C*).

Two studies carried out showed the involvement of B cells in the development of del5q- syndrome (Jaju et al., 2000, Nilsson et al., 2000). One of the three cases of 5q- syndrome in the study conducted by *Jaju et al* showed involvement of B cell. Work undertaken by *Nilsson et al* demonstrated that one out of nine patients in his cohort had B-cell involvement in all the patients, a minimum of 94% of cells in undifferentiated CD34+CD38- were 5q- and in 3 of 5 patients 5q- aberrations were found in a large fraction of $CD34^{+}CD19^{+}$ proB cells suggesting that a lympho myeloid is the primary target in 5q- deletions and that 5q- represent an early event in MDS pathogenesis.

1.2.3 Chromosome 7 and 7q deletions

Monosomy 7 or interstitial loss of 7q, unlike isolated 5q- is associated with poor prognosis. Either in isolation or part of a complex karyotype, about 10% of patients carry abnormality of chromosome 7 and this frequency increases to 50% in tMDS patients having a history of treatment with alkylating agents (Christiansen et al., 2004). Several commonly deleted regions have been mapped for chromosome 7q deletions (7q21, 7q22, 7q32-33, 7q35-36) but the underlying molecular lesions have not been characterised (Asou et al., 2009, Le Beau et al., 1996). *EZH2* gene located on chromosome 7q36 is found to be mutated in patients with -7/7q deletions and alter the epigenetic state of haematopoietic stem cells. Conditional heterozygous deletion of a 2.5 kb region in murine hematopoietic stem cells to a region syntenic to human 7q22 showed no phenotype indicating that this region does not occupy a tumor suppressor gene (Wong et al., 2010).

1.2.4 Trisomy 8

Present as a sole abnormality in approximately 8% of the patients, trisomy 8 is the only recurrent observed chromosomal amplification. It is seen as an intermediate cytogenetic abnormality. The life span of patients with +8 is less than half the median expected survival of patients with normal karyotype. However young patients with refractory anaemia of short duration, having a sole +8 abnormality and carrying *HLADR15* respond well to immunotherapy (Sloand et al., 2008). Patients with +8 MDS often have expansion of Vβ restricted $CD8^+$ T cells which returns to a normal polyclonal distribution after immunosuppressant therapy but bone marrow examination shows persistence and even expansion of +8 clone suggesting that the immune system targets diseased cells and as a consequence impairs normal

haemopoiesis. Trisomy 8 cells show high level of antiapoptotic genes and are resistant to irradiation compared to normal cells (Lim et al., 2007).

1.2.5 Other Chromosomal Abnormalities

Chromosomal lesions like del Y, 20q-, 3q26 have been found in MDS. Both –Y and 20q are considered to be in the same favourable cytogenetic risk group as patients with a normal karyotype. There is no relationship of loss of chromosome Y with disease pathogenesis while interstitial loss of 20q does appear to be pathogenic not only in MDS but also myeloproliferative disorders and AML (Wiktor et al., 2000). 19 genes have been deduced from the 20q CDR of patients with MDS and AML but none of them is involved in pathogenesis of disease (Bench et al., 2000, Wang et al., 2000). Both of the chromosomal lesions are useful for confirming the presence of clonal haemopoiesis and response to treatment but neither lesions are sufficient to make the diagnosis. Recurrent translocation and inversions of 3q26 have been associated with poor prognosis in AML and rare cases of MDS. Breakpoints typically include MDS1-EVI1 (MECOM) gene locus (Poppe et al., 2006).

1.2.6 SNP (Single Nucleotide Polymorphism) Karyotyping

Various DNA based array technologies have been introduced to facilitate the distinction between a normal and malignant genome. Although metaphase cytogenetics has been the gold standard and has been reliably used in clinical karyotyping, it has its limitations. It is time consuming and technically demanding. Its sensitivity is low as it depends on the proportion of clonal cells in the tested samples and resolution depends on the location of the lesion with regards to banding pattern. The need for cellular proliferation to obtain metaphase spreads is a limitation because a significant number shows no cell division. Most importantly

metaphase cytogenetics (MC) cannot detect Acquired somatic Uniparental Disomy (AS-UPD) (Tiu et al., 2011). High resolution genome wide techniques such as SNP genotyping arrays can detect chromosomal microdeletions and areas of copy neutral loss of heterozygosity (CN-LOH) also known as acquired uniparental disomy (UPD). The first study using SNP arrays in a cohort of 119 low risk MDS was performed by *Mohamedali et al* in 2007. The study showed greater than 2 Mb of UPD in 46% of patients with the most prominent UPD found on chromosome 4 (Mohamedali et al., 2007). A multicentre international collaboration was formed to study the clinical applicability and prognostic significance of combining MC and SNP-A in a large cohort of 430 patients with MDS (n=250), MDS/MPN (n=95) and sAML (n=85). Combining SNP-A with MC, chromosomal lesions were identified in 74% of patients as compared to 44% with MC alone. Survival outcomes (Overall survival, progression free survival and event free survival) were poor in patients whose chromosomal lesions were detected by combining SNP-A and MC than with MC alone. A similar observation in survival outcomes was seen in patients with low risk disease when classified according to WHO criteria. Worse overall survival was also found in low risk patients with additional SNP defects than with patients with no additional defects when classified according to the International Prognostic Scoring System (IPSS) (Tiu et al., 2011). Many novel gene mutations have been detected by SNP array technology because regions having UPD are associated with genes that are often mutated which include *TET2* in MDS/MPN, *CBL* in CMML, *EZH2* in MDS, CMML and AML.

1.3 Genetic Mutations

Tumorigenesis is a multistep process that involves irreversible genetic alterations in the DNA sequence such as mutation, deletion and translocations that contribute to progressive transformation of normal cells to malignant cells (Fig 1.2). The term epigenetics refers to the heritable component of cellular phenotype that are not mediated by changes in the genomic DNA sequence (Bejar et al., 2011a) Methylation of cytosine residues and covalent modification of histones are the most relevant molecular mediators of the epigenetic state.

DNA methylation at cytosine that precedes a guanosine in CpG dinucleotide represents a crucial epigenetic modification. Methylation occurs at the 5C position of pyrimidine ring and is catalysed by DNA methyl transferases (DNMTs) resulting in the formation of 5-methyl cytosine. If the deamination of cytosine to form the base thymidine is not recognised and repaired, a cytosine to thymidine change remains resulting in depletion of CpG dinucleotides. However small (0.5kb to several kb) stretches of CpG rich regions termed CpG islands are found in the promoter regions of approximately half of the genes and are unmethylated, that does not interfere with gene expression. The exception to the unmethylated states of CpG island are silenced gene alleles coding for imprinted genes and genes encompassed within regions of X chromosome inactivation.

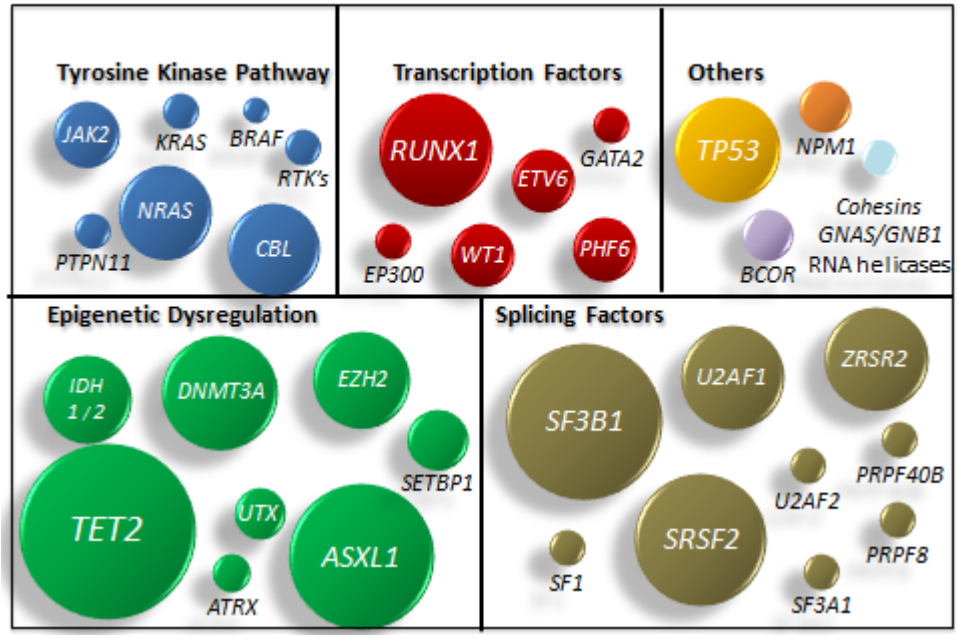


Figure 1.2: Point mutations in MDS. Percentage of mutations in genes found in different pathways that are involved in the pathogenesis of MDS. A pair of genes could show a tendency of either cooccurence or mutual exclusivity. Functional redundancy is implied by mutually exclusive gene pair especially if they occur in the same pathway. Members of the splicing machinery occur exclusive to one another; mutual exclusivity is also seen between *TET2* and *IDH2* although both of them are linked to disordered hypomethylation. On the other hand mutations in *DNMT3* are found to co-occur with *SF3B1* and of *ASXL1* with *EZH1*.

The fundamental repeating unit of chromatin is the nucleosome that consists of 146 bp of DNA wrapped around an octamer of core histone proteins H2A, H2B, H3 and H4. The histone tails are subjected to post translational modification that play an important role in controlling and maintaining chromatin structure and are also the sites of epigenetic modification. Acetylation of histones by histone acetyl transferases (HAT) is associated with nucleosome modelling and transcriptional activation while deacetylation by histone deacetylases (HDAC) represents chromatin condensation and transcriptional repression. Another modification of histones is methylation carried by various histone methyl transferases, which can act as marker of activation as well as repression.Two critical methylation events such as H3K9 and H3K27 methylation on the N-terminus of H3 are markers of

transcription silencing and inactivation, whilst methylation of H3K4 is found near the open chromatin structures of active genes (Galm et al., 2006).

1.3.1 Epigenetic modifiers gene mutations

EZH2 (Enhancer of zeste homolog 2) is a gene involved in epigenetic regulation and is mutated in 6% of MDS cases and over 10% of MDS/MPN overlap. *EZH2* gene is located on Chr. 7q36.1 and codes for a histone methyl transferase that forms the catalytic subunit of the polycomb repressor complex 2 (PRC2). di and trimethylation of lysine 27 on histone 3 catalysed by the PRC2 complex is an epigenetic marker for transcriptional repression. In solid tumors, high expression of *EZH2* is associated with aggressive disease and worst outcomes. *EZH2* is often mutated in myeloid malignancies resulting in the loss of the catalytic subunit and loss of PRC2 function leading to an increase in stem cell renewal and differentiation. Truncated mutations occur throughout the gene while missense mutations are usually clustered in the C-terminal catalytic SET (suppressor of variegation, enhancer of zeste, trithorax) domain, and also in the cysteine-rich and SANT (switch-defective protein 3, adaptor 2, nuclear receptor corepressor, transcription factor) domains. The mutations are hemizygous (loss of one allele) or biallelic (via microdeletions of 7q or UPD) in most cases (Makishima et al., 2010, Nikoloski et al., 2010). *EZH2* Y641 codon is frequently mutated in follicular and diffuse large B-cell lymphomas, but not in MDS. *In vitro* studies have shown that *EZH2* mutations found in MDS patients results in loss of function (leading to reduced H3K27 trimethylation). *EZH2* mutations are associated with poor survival. The H3K27 methylation mark recruits DNMT3 to site of *de novo* methylation providing a link between the two repressive epigenetic pathways (Ernst et al., 2010).

Mutations in additional sex-comb like-1' *(ASXL1*) gene located on chromosome 20q11.1 have been described in roughly 10% of MDS and MPN, 17% of AML and >40% of patients with CMML (Boultwood et al., 2010, Gelsi-Boyer et al., 2009). *ASXL1* belongs to a three-member family of enhancers of trithorax and polycomb proteins (ASXL1, 2, 3). It encodes a chromatin binding protein involved in epigenetic regulation and is responsible for recruiting polycomb and trithorax complexes to specific loci. *ASXL1* protein has two domains, a plant homeo domain finger and nuclear receptor box domain and functions as a ligand dependent coactivator of retinoic acid receptor. It has been shown to mediate its effect through direct interaction either with histone acyl transferases encoded by *NCOA1* or histone demethylase encoded by *LSD1* (Carbuccia et al., 2009). *ASXL1* mutations are thought to promote myeloid transformation through loss of PRC2 mediated myeloid repression (Abdel-Wahab et al., 2012). The vast majority of *ASXL1* mutations found in myeloid malignancies affect the glycine rich exon 12 and are frameshift and nonsense mutations resulting in C-terminal truncation of protein and loss of C terminal Plant-homeo-domain (PHD) finger. Truncated protein is unable to bind methylated histone lysine or interact with chromatin modifiers (Acquaviva et al., 2010).

Progressive multilineage cytopenias and dysplasia with increased numbers of haematopoietic stem/progenitor cells characteristics of human MDS is seen in mice with haematopoietic deletion of Asxl1. Most of the *ASXL1* mutations in myeloid patients are seen in its C terminal, resulting in a deletion of PHD protein interacting domain creating dominant negative protein that inhibits its wildtype counterpart as well as other members of the polycomb multiprotein complex.

Ten-eleven translocation gene (*TET1*) was named as a fusion partner with the mixed lineage leukaemia (MLL) gene in t(10;11)(p12;q23) AML (Lorsbach et al., 2003). It has two other paralogs: *TET2* and *TET3*. *TET2* encodes a 2-oxoglutarate and Fe (II) dependent deoxygenase that catalyses the conversion of 5-methyl cytosine (5mc) to 5-hydroxy methyl cytosine (5hmc). TET enzymes can further convert (5hmc) into 5-formycytosine (5fC) and 5-carboxyl cytosine (5caC) (Fig 1.3). The N-terminal consists of an alpha helix followed by continuous sheets of β strands, typical of the double-stranded β helix (DSBH) folds of the 2OG-Fe(II) oxygenases. *TET1* and *TET3* also possesses a CXXC domain, a binuclear zinc-chelating domain that is found in chromatin associated protein.

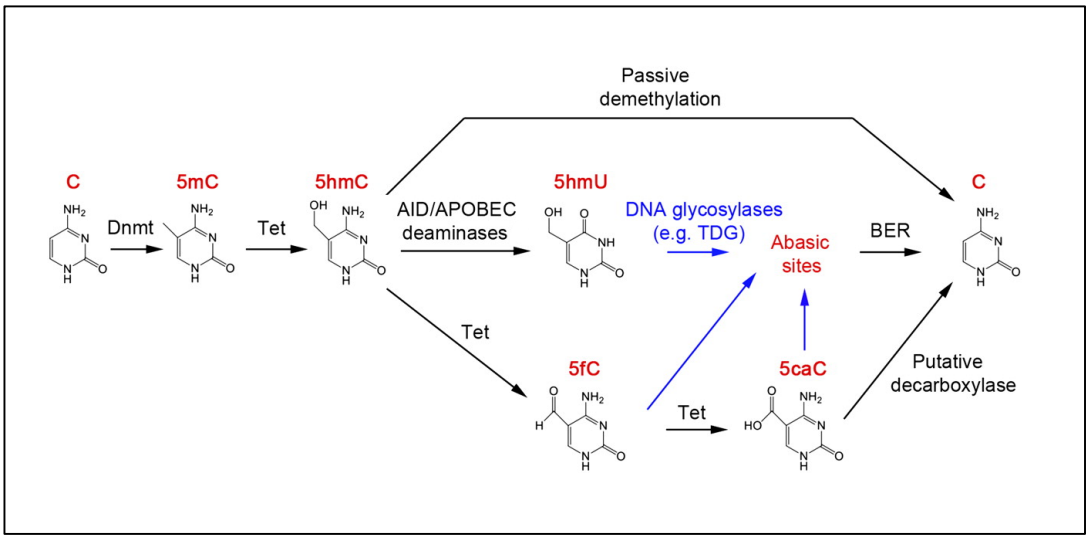


Figure 1.3: DNA methyl transferases (Dnmt) methylate cytosine to 5methyl cytosine (5mc). 5mc can be hydroxylated to 5 hydroxy methyl cytosine (5hmC) by Tet family of deoxygenases. Process of replication dependent passive demethylation can convert 5hmC to cytosine due to its poor recognition by Dnmt1. 5hmC can be further oxidized to 5 formyl cytosine (5fC) and 5 carboxy cytosine (5caC) by Tet proteins. 5hmC can be deaminated to 5 hydroxy methyl uracil (5hmU) by AID/APOBEC deaminases. Further TDG (thymine DNA glycosylase) a DNA glycosylases excises 5hmC, 5fC and 5caC from DNA.

TET2 gene located on chromosome 4q.24 is the most commonly mutated gene in MDS. It is found mutated in 20% MDS, 10% of MPN, 30-50% of CMML and 25% of secondary AML (Delhommeau et al., 2009, Kosmider et al., 2009, Mullighan, 2009), *TET2* mutations are generally restricted to the C terminal end of the protein and result in loss of function. Increase in the level of 5mc and reduction in level of 5hmc is observed with *TET* mutations leading to DNA hypermethylation and gene silencing (Ko et al., 2010, Ponnaluri et al., 2013, Tang et al., 2009). *Crusio et al* showed that reduced *TET2* expression leads to enhanced stem cell function and also augments transformation to AML (Moran-Crusio et al., 2011). Decreased expression of *TET2* has also been found in progenitors, erythoid precursors and granulocytes in patients with no detectable mutations suggesting alternative mechanism of *TET2* deregulation and highlighting the physiological and tumor suppressor function of this protein (Langemeijer et al., 2009).

DNA methyl transferases (DNMTs) mediate the methylation at 5'position of cytosine residues in a CpG dinucleotide. DNA methylation is performed by three different methylases de novo methylation by *DNMT3A* and *DNMT3B* and maintenance of methylation by *DNMT3*. Mutation in *DNMT3A* was first identified in an AML patient with normal karyotype and was subsequently found in 8% of *de novo* MDS. Heterozygous missense mutation that converts arginine to histidine at position 882 (R882H) is the most frequent *DNMT3A* mutation in AML and MDS, which reduces the catalytic activity *in vitro* although it has not been studied in combination with wildtype allele (Yamashita et al., 2010). Compound heterozygote frameshift and nonsense *DNMT3A* mutations do occur suggesting that loss of function mutations exist. Loss of dnmt3a$^{-/-}$ leads to a competitive growth advantage over wildtype in murine dnmt3a$^{-/-}$ hematopoietic stem cells (Challen et al., 2011).

Another recurrently mutated gene found in haematological malignancies are Tri carboxylic acid cycle (TCA) enzymes. *IDH1* and *IDH2* were identified as mutated oncogenes in a high percentage of glioma and secondary glioma patients (Yan et al., 2009). Whole genome sequencing of an AML patient sample led to discovery of *IDH1* mutation in 8% of AML samples (Yoshida et al., 2011). *IDH1* and *IDH2* mutations have been confirmed in AML, in MPN at the time of leukaemic transformation and in rare cases of MDS (Abdel-Wahab et al., 2010, Kosmider et al., 2010, Paschka et al., 2010, Tefferi et al., 2010, Thol et al., 2010).

IDH1 and *IDH2* are homodimeric nicotinamide adenine dinucleotide phosphate ($NADP^+$) dependent enzymes that convert isocitrate to α-Ketoglutarate (αKG) and reside in cytoplasm and mitochondria respectively. Heterozygous mutations alter the functions of the encoded enzymes producing 2-hydroxyglutarate (2HG) instead of α-KG which inhibits the functions of many enzymes such as TET family members, histone demethylases, and prolyl hydroxylases (Losman and Kaelin, 2013) (Fig 1.4). Most of the mutations are missense mutations of conserved codons (R132 for *IDH1* and R140 and R172 for *IDH2*) and no frameshift or early truncations have been identified. Mutations of *TET2* and *IDH* gene appear largely exclusive of each other in MDS and AML suggesting that they engage similar oncogenic mechanisms of epigenetic regulation (Figueroa et al., 2010).

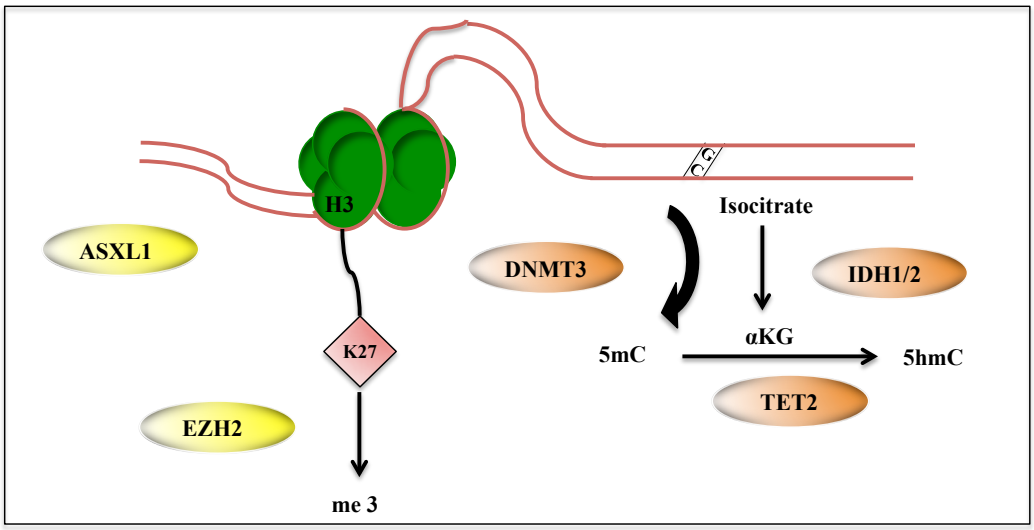


Figure 1.4: Pathways in MDS encoding epigenetic modifiers and transcriptional factors. Trimethylation of lysine 27 on carboxyl terminal tail of histone3 a nucleosome component is associated with transcriptional repression. *ASXL1* and *EZH2* are members of polycomb complex protein that maintain the repressive state. *DNMT3* is a methyl transferase that converts cytosine in CpG islands to 5 methyl cytosine (5mC) which is associated with repression. 5mC is converted to 5hmC by *TET2* in presence of α ketoglutarate (αKG) produced by IDH.

1.3.2 Mutations in Transcription regulation and cell signalling pathway

TP53

TP53 is a tumour suppressor gene located on chromosome 17p13 and its gene product p53 is activated by a variety of cellular stresses instigating cell cycle arrest, apoptosis and a DNA repair response. Mutations in *TP53* although not associated with a specific morphological or clinical phenotype are associated with adverse disease features. In MDS mutations in *TP53* have been described in high risk MDS particularly with complex karyotype and therapy-related MDS. Mutations in *TP53* gene occur in 10% of MDS with nearly a third to half of the patients having complex karyotype and are associated with poor prognosis after adjusting the IPSS subgroup (Bejar et al., 2011b, Kulasekararaj et al., 2013). Mutations in *TP53* or p53 protein expression in del(5q) setting predict poor response to lenalidomide and higher progression to AML (Jadersten et al., 2011, Saft et al., 2014).

Activation of p53 is important for the erythopoietic defects associated with haploinsufficiency for *RPS14* (Ebert et al., 2008). *Barlow et al* created a mouse model for 5q- syndrome and showed that segmental haploidy of a region between *Cd74* and *Nid67* syntenic to a region within the CDR of 5q-syndrome containing the gene *RPS14* resulted in macrocytic anaemia a key feature of 5q- syndrome. The bone marrow of these mice showed a prominent dysplasia within the erythroid and monolobulated megakaryocyte lineage, decrease in erythroid and myeloid progenitor cells which was accompanied by an increase in p53 positive cells and increase apoptosis of bone marrow cells proving the evidence that haploinsufficiency of *RPS14* in 5q- causes ribosome disorganised biogenesis. Deletion of p53 in *Cd74*/*Nid67* deficient mice ameliorated thrombocytopenia and reduction in number of progenitor cells supporting the evidence that ribosomal stress leads to activation of p53 (Barlow et al., 2010). A leading hypothesis suggests that ribosomal haploinsufficiency leads to accumulation of free ribosomal proteins due to disrupted ribosomal biogenesis that binds to MDM2 a negative regulator of p53 which results in activation of p53 leading to apoptosis and cell cycle arrest (Dutt et al., 2011).

RUNX1

*RUNX1 (*also known as *AML1 or CBFA2)* is the second most commonly mutated gene in MDS and was identified as a translocation partner to *ETO* (now called *RUNX1T1*) in cases of t(8:21) in AML. It encodes for one of the two heterodimeric subunits of the human core transcription factor (CBF). *RUNX1* translocations are more common in AML as compared to MDS, however point mutations are found in *de novo* MDS (7 -15%) and at a higher frequency in therapy related MDS. In both

MDS and AML, *RUNX1* mutations are markers of poor prognosis (Chen et al., 2007, Steensma et al., 2005).

A proximal Runt homology domain (RHD) essential for both DNA binding and distal transactivation domain (TAD) essential for protein protein interaction and recruitment of cofactors is found in RUNX1 protein. Missense mutations are found in the Runt domain, whilst nonsense and frameshift mutations are found throughout the length of the protein and disrupt the transactivation domain (Fig 1.5). This distinction may be physiologically relevant since Runt domain mutations with impaired DNA binding can function in a dominant negative manner. RUNX1 activity will be reduced by these mutations as compared to loss of function mutation by one allele.

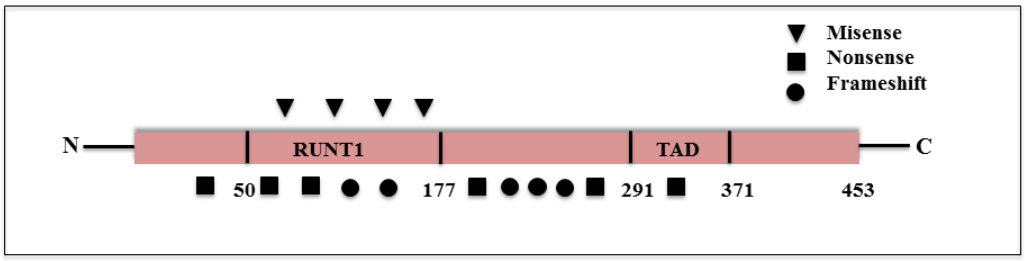


Figure 1.5: Localisation of various domains found in RUNX1 protein.The RUNX1 protein has two domains the proximal RUNT1 domain responsible for DNA binding and the distal transactivation domain (TAD) responsible for protein protein interaction. Missense mutations are found in the Runt domain while nonsense and frameshift mutations are found throughout the length of the protein.

Germ line mutations in *RUNX1* cause familial platelet disorder a rare human disease with an inherent susceptibility to develop acute myeloid leukaemia suggesting that this gene may act as a tumor suppressor gene in this setting (Song et al., 1999). *Runx1* knockout mice die without achieving complete haematopoiesis few days after birth (Dowdy et al., 2010). Excision of *Runx1* in the adult haematopoietic compartment of these mice leads to extramedullary haematopoiesis,

lymphoid defects, expansion of myeloid progenitor pool and inefficient platelet production but do not succumb to AML (Growney et al., 2005).

RAS

The RAS proto-oncogene belongs to small GTPase family and exists in three different isoforms *N-RAS*, *K-RAS* and *H-RAS*. In MDS *NRAS* the most common mutation is present in 10-15% of patients while 1-2% having *KRAS* mutations. Considered in isolation these mutations are associated with poor prognosis and progression to AML. RAS proteins are GTPases that act as signal switch molecules in controlling signalling pathways involved in cell proliferation and cell survival. Ras proteins are normally tightly regulated by guanine nucleotide exchange factors (GEFs) promoting GDP dissociation and GTP binding and GTPase-activating proteins (GAPs) that stimulate the intrinsic GTPase activity of Ras to switch off signalling. The RAS signalling pathway is constitutively activated by RAS mutations in exon 12, 13 and 61 resulting in an increase in intracellular RAS GTP which in turn activates various pathways like the BRAF mitogen- activated and extracellular signal-regulated kinase (MEK), extracellular signal-regulated kinase (ERK) cascade which determines proliferation and becomes deregulated in certain cancers. (Schubbert et al., 2007).

C.CBL

The CBL gene is located on chromosome 11q23.3 and its gene product is a tyrosine kinase associated ubiquitin ligase that negatively regulates signalling through these receptors by targeting them for degradation. In CMML and in patients with juvenile myelomonocytic leukaemia it is mutated in 15% of cases, but in MDS it is found in fewer than 5% of patients. CBL mutations result in increased receptor tyrosine kinase (TK) levels and phosphorylation of STAT5, which are believed to mediate

hypersensitivity of these mutant cells to a wide variety of growth factor and cytokines (Sanada et al., 2009, Sargin et al., 2007). *CBL* mutations are biallelic, which suggests that CBL is a tumour suppressor gene and loss of wildtype is advantageous. However most of the *CBL* mutations are localised in the N- terminal region in exons 8 and 9 leaving the rest of the protein intact suggesting a gain of function. Frameshift and early truncation mutations are uncommon. The common CBL mutations are known to encode a dominant negative protein that inhibits the function of the wildtype and its homolog CBLB (Fig 1.6).

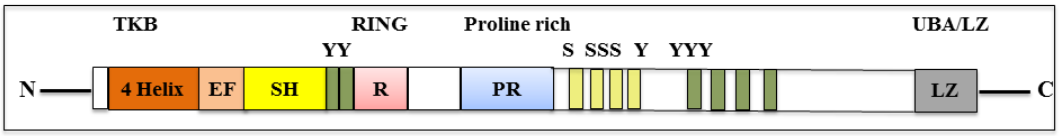


Figure 1.6: Schematic representation of the domain architecture of C.CBL protein.The C.CBL consists of a highly conserved N-terminal region that is comprised of a tyrosine kinase binding domain (TKB) consists of 4 helix, a calcium binding EF domain, a variant SH2 domain linked with the RING finger domain. The C-terminal consists of a proline rich stretches multiple serine and phosphorylated tyrosine sites and an ubiquitin associated UBA domain and leucine zipper.

Modest cytokine hypersensitivity is seen in mouse cells deficient for CBL that is greatly enhanced by the introduction of a mutant *CBL* gene perhaps due to the inhibition of partially compensatory CBLB activity (Rathinam et al., 2008). In MDS *CBL* mutations are found to be exclusive of several other commonly mutated genes including *FLT3*, *KIT NPM1*, *CEBPA*, *PTPN11* and *NRAS*.

JAK2

Activating mutations of *JAK2* (Janus Kinase 2) form a part of diagnostic criteria for polycythaemia vera and are found in roughly half of the patients with thrombocythemia and primary myelofibrosis. JAK2 is a nonreceptor tyrosine kinase that transduces the signal from several cytokine receptors including the erythropoietin (Epo), G-CSF, and thrombopoietin (Tpo) receptors. Depending on

the receptor to which it is attached it can activate a large number of signalling pathway involved in cytokine mediated cell survival, proliferation and differentiation, Only the common mutation V617F has been reported in MDS and is found in 5% of unselected patients and does not have any prognostic significance (Steensma and Tefferi, 2008). However, MDS/MPN (Myeloproliferative neoplasm) crossover syndrome RARS-T is an exception where *JAK2V617F* is found in 50% of patients. Cases of wildtype *JAK2* RARS have been defined in which thrombocythemia occurs only at the time of acquisition of V617F mutation (Malcovati et al., 2009). Given the high rate of *JAK2* mutation in MDS/MPN there appears to be a cooperating advantage between those abnormalities responsible for the ring sideroblast phenotype and the constitutive activation of JAK2. The nature of this interaction is not yet understood.

1.3.3 Splicing mutations

Whole genome sequencing identified mutations in a completely new class of genes involved in mRNA splicing. *SF3B1* was the first gene to be identified, which was particularly mutated in 68% of cases with MDS subtype refractory anaemia with ringed sideroblast (RARS) (Papaemmanuil et al., 2011). All the mutations appeared to be heterozygous mutations and were clustered from exon 12 to 15 of the gene. K700E accounted for 59% of the variants observed. Several other hotspots for amino acid mutations E622, R625, H622, K666, I704 were also observed. All the mutations identified were found in the fourth, fifth, and sixth HEAT domains localised in the C-terminal. Transcriptome profiling suggested that *SF3B1* mutation was associated with downregulation of essential mitochondrial gene network. Further patients with *SF3B1* had overall longer survival and no deleterious effects

than patients without *SF3B1* mutations. Since *SF3B1* mutations could readily be identified in peripheral blood as well as ring sideroblast in bone marrow the study proposed the use of *SF3B1* as a screening for patients for MDS. It is also found to be recurrently mutated in chronic lymphocytic leukaemia (CLL) and at low frequencies in solid tumors (Wang et al., 2011). In a parallel study *Yoshida et al* identified mutations in multiple components of the RNA splicing machinery, *SF3B1*, *SRSF2*, *U2AF1* (*U2AF35*), *ZRSR2*, using whole exome sequencing in 57% of MDS patients. Most of the mutations were mutually exclusive and affected genes involved in 3' splice site recognition step of pre mRNA processing inducing abnormal RNA splicing and haemopoiesis. In case of *U2AF35* mutations were found clustered at two highly conserved amino acids positions S34 or Q157, *SRSF2* mutations exclusively occurred at position P95 while most of the *SF3B1* mutations were found at position at K700 and to a lesser extend involved K666, H662 and E622 while in case of *ZRSR2* they were widely distributed throughout the region. Gene set enrichment assay (GSEA) analysis disclosed an enrichment of genes involved in nonsense mediated decay (NMD) pathway among the significantly up regulated genes in mutant *U2AF35*-transduced HeLa cells. Most of the mutations identified in *SF3B1*, *SRSF2* and *U2AF1* to date are missense mutations affecting invariant positions suggesting a gain of function or change of function. In the case of *ZRSR2* missense nonsense and frameshift mutations have been reported suggesting loss of function (Papaemmanuil et al., 2013, Thol et al., 2012). Splicesome mutations have been found to co-occur with mutations in epigenetic modifiers in MDS. Mutations in *SF3B1* has been found to co-occur with mutations in *DNMT3A* the methyl transferase, *TET2* mutations co-occur with mutations in *SRSF2* and *ZRSR2* and *U2AF1* has been associated with *ASXL1* or *TET2*.

Splicing factor mutations were found in 38% of cases in a cohort of 154 MDS patients in a study carried out by *Mian et al*. *SF3B1* mutations were found in 80% of cases with RARS /RCMD/RS and showed beneficial prognostic impact on overall survival. Mutations in *SRSF2*/ *U2AF1* occurred in advance forms of diseases like RAEB and CMML and coexisted with mutations in *NRAS*, *FLT3* and *RUNX1* with known oncogenic mutations (Mian et al., 2013). Subsequently the same group showed that *SF3B1* mutations arise from haematopoietic stem cell compartment thereby providing a therapeutic target (Mian et al., 2015).

1.3.4 Cohesin mutations

The cohesin complex which is involved in holding the sister chromatids together, post replicative DNA repair, transcriptional regulation is found to be mutated in MDS and other myeloid malignancies. Cohesin mutations actually fall into two categories, truncation and frameshift mutations reported in *STAG2* and *RAD21* and missense mutations reported in *SMC1A* and *SMC3*. While most evidence for cohesin mutations in myeloid leukaemia has come from AML, cohesin mutations have also been found in related myeloid malignancies. In 2010, a comparative genomic hybridisation identified loss of one copy of *RAD21* in one patient with chronic myelomonocytic leukaemia and complete loss of *STAG2* in another patient with *de novo* acute myeloid leukaemia in 167 myeloid disease samples. These initial studies provided important hints that mutations in cohesin complex have a role to play in tumorigenesis (Rocquain et al., 2010). The first report of somatic cohesin mutations in AML was identified in *SMC3* gene in a cohort of 200 patients at a frequency of 3% (Ding et al., 2012). Subsequently this result was confirmed and extended by TCGA (The Cancer Genome Atlas) identifying mutations in each of the cohesin subunit (*SMC1A*, *SMC3*, *RAD21* and *STAG2*) at a frequency of ~3%

and a cumulative cohesin mutations frequency of 13% (13/200) in adult *denovo* AML. All the mutations were mutually exclusive of one another. *Kon et al* (Kon et al., 2013) and *Thota et al* (Thota et al., 2014) determined the frequency of cohesin mutations in both *de novo* and secondary AML (as well as in other types of myeloid malignancies). In both the studies AML harboured the most frequent mutations in core cohesin subunits. *Kon et al* identified mutations and deletions of the cohesin subunits in 13% (16/120) of *de novo* AML and 8% (3/37) of secondary AML in a mostly mutually exclusive manner. In contrast *Thota et al* identified a higher frequency of cohesin mutations in secondary AML samples (20%; 30/149) than in *de novo* AML (11%; 32/301). In both the studies most of the mutations were identified in *STAG2* followed by *RAD21* and then *SMC3*. Further evidence of cohesin's involvement in myeloid malignancies came from the work of Haferlach showing that approximately 15% of patients with MDS harbour mutations in cohesin complex. *STAG2* and *SMC1A* were associated with poor survival outcomes that strongly suggested that cohesin mutations were central to development and prognosis of MDS (Haferlach et al., 2014)(Walter et al., 2012). Cohesin mutations have also been associated with pathogenesis of DS-AMKL (Down syndrome acute megakaryoblastic leukaemia). Genome profiling and whole exome sequencing in a cohort of Transient abnormal myelopoiesis (TAM), DS-AMKL and non-DS-AMKL revealed that DS-AMKL emerged from a pre-existing TAM clone (already having an underlying constitutive trisomy 21 and *GATA1* mutation) after acquisition of additional mutations with major mutational targets (53%) in members of the cohesin complex (Yoshida et al., 2013).

One of the members of the cohesin complex *STAG2* located on X chromosome has been found to be mutated in a large variety of cancers. *Solomon et al* reported

STAG2 to be mutated in 21% Ewing's sarcoma and 19% of melanoma and glioblastoma. Loss of *STAG2* in this study, however, was associated with chromosomal instability (Solomon et al., 2011). In another study, truncated mutations in *STAG2* in 36% of papillary non-invasive urothelial carcinoma and 16% of invasive urothelial carcinomas of bladder were detected. Tumours with *STAG2* mutations were also shown to have concurrent p53 overexpression or mutations (Solomon et al., 2013). The highest incidence of *STAG2* mutations has been reported in bladder cancer (24.7%) (Taylor et al., 2013). Most of the mutations in this study were inactivating mutations (missense, frameshift, splicing) leading to loss of protein function. Interestingly, the mutations in this study were not associated with chromosomal instability, such as loss of whole chromosomes (aneuploidy) (Taylor et al., 2013). In another study, *Martinez et al* predicted *STAG2* as a tumour suppressor gene that was found to be mutated in 12 of 77 urinary bladder tumors (15.6%). Similarly to previous studies, loss of *STAG2* was not associated with aneuploidy and had a better prognosis in both NMIBC (Non muscle invasive bladder cancer) and MIBC (muscle invasive bladder cancer) (Balbas-Martinez et al., 2013). Loss of expression of *STAG2* has also been found in 27% gastric cancers, 23% of colorectal cancers (CRC) and 30% prostrate carcinomas (PCA) (Kim et al., 2012). *STAG2* inactivation cooperated with *KRAS* mutation as an early event in the progression and evolution of human pancreatic ductal adenocarcinoma (PDA) (Evers et al., 2014). Mutual coexistence of *STAG2* and *TP53* was found to be associated with worst prognosis in a study to reveal secondary mutations that contribute to disease progression. Data from this study suggested that *STAG2* and *TP53* co-operate towards genomic instability in an

aggressive disease like Ewing's sarcoma. Mutual exclusivity was also found between *STAG2* and *CDKN2A* lesions (Tirode et al., 2014).

The pathophysiology of MDS and its progression to AML involves cytogenetic, epigenetic and genetic aberrations. Data arising from whole genome sequencing (WGS) have shown that the progression of MDS to AML is complex. Further second generation sequencing have shown more tantalizing results like the occurance of small clones of MDS related genes in individuals who are either healthy or have mild cytopenias with no definitive signs of MDS. Better understanding of the molecular landscape of MDS has important clinical implications. The discovery of mutations in members of the cohesin complex has opened a new field that needs to be unravelled in the pathogenesis of MDS so a part of my thesis will focus on the occurrence and clinical correlation of cohesin mutations in MDS patients.

1.4 Cohesin Complex

The chromosomes are at the very heart of all genetic processes. Accurate chromosomal segregation at mitosis is essential for the maintenance of genomic integrity. Sister chromatids need to be tightly coupled from the time of generation in S phase until metaphase, so as to undergo appropriate segregation at anaphase. The cohesin complex plays a pivotal role in the establishment of the cohesion between sister chromatids both during mitosis as well as meiosis. Cohesin belongs to the conserved group of protein complexes that are dedicated to chromosome biology. It has now been recognised that the cohesin complex also plays pivotal roles in maintaining the genomic integrity of the cell by holding sister chromatids together during post replicative DNA repair, and through its regulation in gene transcription.

Defects in the cohesin complex have been associated with severe developmental diseases collectively termed as “cohesinopathies” examples of which include Cornelia de Lange Syndrome (CdLS), Roberts Syndrome (RBS)/ SC phocomelia (SC) and Warsaw breakage syndrome (Mannini and Musio, 2011).

1.4.1 Architecture of Cohesin complex

The cohesin complex is composed of four subunits, SMC1, SMC3, (RAD21, Mcd1, Scc1) kleisin and stromalin (SA, Scc3, STAG1/2) proteins which are conserved across the species from yeasts to mammals. The structural organisation of individual components have been well characterized in budding yeast and the complex shows structural homology with cohesin from other species with sequence conservation among the subunits, as well as biochemical and electron microscopic similarities (Dorsett, 2011).

Smc1 and Smc3 are members of the structural maintenance of chromosome (SMC) group of chromosomal ATPases that are conserved from bacteria to humans. Most prokaryotes have only one Smc protein which form a homodimer, while there are six Smc proteins in eukaryotes (Smc 1-6) which form a heterodimer. Each Smc subunit is self-folded by anti-parallel coiled coil interactions, so that the N and the C terminal come together creating a rod shaped molecule with an ATP binding cassette (ABC)-like head domain at one end and hinge domain at the other end. Each head domain contains a N-terminal walker A box and a C-terminal walker B box (Ivanov and Nasmyth, 2005, 2007).

Although various models have been proposed as to how the cohesin holds the sister chromatids together, such as ring (embrace model), the two ring model, multimeric bracelet and multimeric rod shaped model, the ring model is accepted widely where by the hinge domain of Smc1 and Smc3 are tightly associated with each other while

the ATPase head domain are physically connected by RAD21 (Scc1, Mcd1) to form a close tripartite ring with an outer diameter of 40nm. ATP binding to the ATPase domain is essential for binding of RAD21/Scc1/Mcd1 with the Smc1/ Smc3 heterodimer (Weitzer et al., 2003). The fourth unit of the complex Scc3 or STAG1/2 binds to RAD21. The sequence of this protein has HEAT repeats that are responsible for protein protein interaction. The cohesin complex consists of various other regulatory subunits like PDS5, WAPL and sororin that play a role in loading and unloading of cohesin (Fig 1.7).

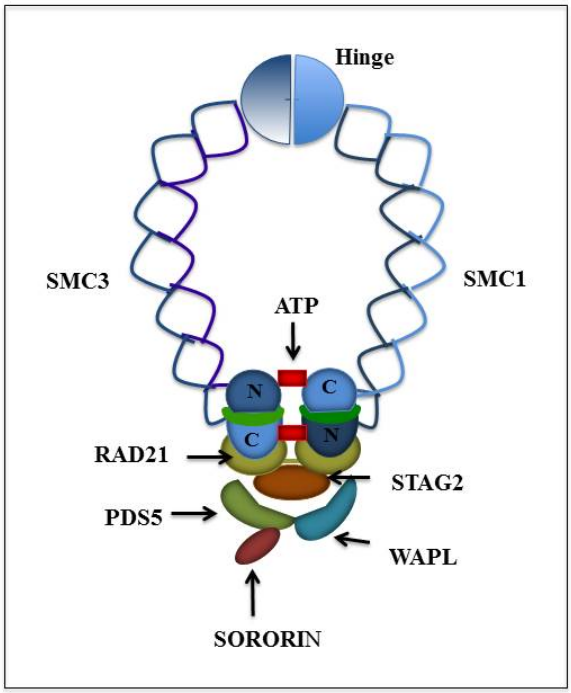


Figure 1.7: Structure of the Cohesin complex. The cohesin complex consists of a heterodimer of SMC1 and SMC3 which fold upon themselves in the hinge region to form antiparallel coiled coil arms with the amino and carboxy termini coming together in head domains that contain ABC-type ATPases. Cohesin forms a ring like structure with RAD21 bringing the SMC head domains together. STAG2 binds to RAD21. Pds5, Wapl and Sororin are other regulatory factors responsible for loading and unloading of cohesin. The cohesin complex has critical roles in sister chromatid cohesion.

1.4.2 Cohesin Loading and Establishment

The cohesin complex is loaded at the end of mitosis in telophase in higher eukaryotes, but at the end of G1/S boundary in *Sacchromyces cerevisiae* (budding

yeast) by a protein complex kollerin consisting of adherin proteins (Scc2, Mis4, NIPL, Nipped B) and Scc4 (Mau-2). Scc2 are HEAT repeat proteins that physically interact with cohesin. Kollerin binds to the chromosome and along with ATP hydrolysis is required by the Smc proteins for topological binding with the sister chromatids. The mechanism by which cohesin is loaded by Scc2-Scc4 on the chromosome is poorly understood. Experiments by *Nasmyth* and co-workers have predicted that a preformed ring associates with the chromatin and that the DNA enters the ring from the hinge region. The energy requirement for ring opening is derived from the bending of the hinge domain towards the ATPase head resulting in ATP hydrolysis (Gruber et al., 2006). In budding yeast (*Sacchromyces cerevisiae*), cohesin is loaded at CAR sites (Cohesin associated regions) that fall within the intergenic regions and are distinct from the Scc2-Scc4 loading sites. Although the notion that cohesin first binds to Scc2-Scc4 sites and then is translocated to CARs site is argued by evidences suggesting that the loading site can lie within the CARs site suggesting a direct deposition of the complex. Orthologs of Scc2 and Scc4 have been found in fission yeast, drosophila and mammals. A major fraction of mouse and human cohesin sites have been found in the vicinity of the transcriptional insulator, CTCF (CCCTC binding factor required for transcriptional repression) with a preference for regions in the vicinity of transcribed genes. How cohesin is loaded at these sites is not clear. Cohesin needs to be relocated from its site of loading to permanent positions and *Nasmyth* and co-workers have reported that ATP hydrolysis responsible for association of cohesin to chromatin is also responsible for its translocation (Hu et al., 2011).

It has also been shown that the association of cohesin with the chromosome is not enough for pairing of sister chromatids. For this chromosome- associated cohesion,

cohesin must attain a "pairing competent state" which is acquired by the acetylation of the Smc3 protein on Lysine^{112} and Lys^{113} by Eco1/Ctf7 (establishment factor) in budding yeast. In mammals two orthologs (Esco1 and Esco2) also regulate cohesion at Lys^{105} and Lys^{106} Reviewed in (Nasmyth, 2011). Eco1/Ctf7 interacts physically with the sliding clamp Proliferating cell nuclear antigen (PCNA) as well as with the clamp loader replication factor C (RF-C) further suggesting its role in linking cohesin to DNA replication (Kenna and Skibbens, 2003). Other replication associated proteins include an "alternative" RF-C component containing Chromosome transmission fidelity protein 8 (Ctf8), Chromosome transmission fidelity protein 18 (Ctf18), sister chromatid cohesion protein (Dcc1), Chromosome transmission fidelity protein4 (Ctf4) that associates with DNA polymerase α and the Chl1 helicase (close homolog of L1) (Bermudez et al., 2003, Hanna et al., 2001, Mayer et al., 2001, Skibbens, 2004). It remains unclear how the establishment of cohesin is coupled to replication at a mechanistic level and *Losada et al* have proposed various models for it as shown in Fig 1.8 (Losada and Hirano, 2005).

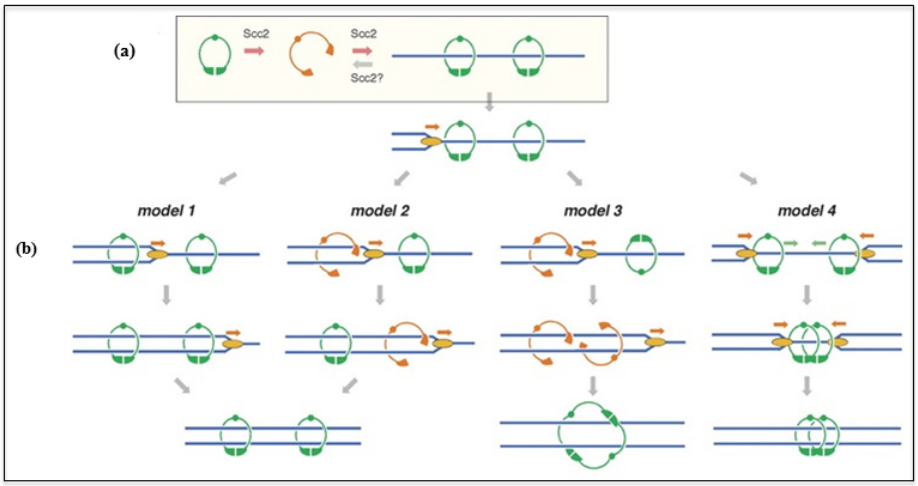


Figure 1.8: Regulation of cohesin dynamics during Interphase. (a) Scc2 promotes loading of cohesin onto the parental chromatid before DNA replication. It may also be involved in transient unloading or mobilization of cohesin. Scc2 could mediate these functions by facilitating disengagement of the SMC head domains of cohesin. Two confirmations of cohesin, with engaged and disengaged head domains, are represented in green and orange respectively. (b) Speculative models for the establishment of cohesion during DNA replication. The replication machinery (yellow oval) could simply pass through the cohesin ring (model 1). Alternatively, passage of the replication fork could impose a conformational change in cohesin. The altered cohesin complex may return to the original conformation (model 2) or may generate a novel cohesive structure (model 3). Model 4 proposes that the replication forks push the cohesin complexes so that they accumulate at termination sites. Replication of the remaining short stretch of DNA could be facilitated by specialized factors such as the alternative RF-C. Subsequent action of the transcriptional machinery could then relocate cohesin from the termination sites to regions of convergent transcription (Losada et al 2003).

The acetylation by Eco1/Ctf7 counteracts the effect of the anti-establishment cohesion factor, releasin formed by Pds5 (BimD or Spo76) and wing associated protein (Wap1, Wpl1, Rad61) and release of Pds5 from the releasin complex. Vertebrate's cells have two Pds5 like proteins (PDS5A and PDS5B) and are associated with maintenance of cohesin. Studies in mouse and human cells have shown that cohesin complexes that include SA1 (cohesin-SA1) and SA2 (cohesin-SA2) mediate sister chromatid cohesion at telomeres and centromeres. PDS5A and PDS5B which can bind to either cohesin-SA1 or cohesin-SA2 both contribute to telomere and arm cohesion whereas PDS5B is specifically required at centromeres

(Losada, 2014). Irr1/Scc3 is the third member of the anti-establishment complex and has been demonstrated as one of the stable components of cohesin complex.

Acetylation also recruits Sororin (Dalmatian) proteins in vertebrates and drosophila that prevents cohesin from disengagement by displacing WAPL from PDS5 although WAPL remains attached to cohesin. Hos1 in yeast and HDAC8 (Human deacetylase 8) in human deacetylates Smc3 in preparation for the next cycle of mitosis (Borges et al., 2010, Nasmyth, 2011).

1.4.3 Regulation of Cohesin

Timely dissolution of cohesin is important for proper segregation of chromosomes. In vertebrates, the bulk of cohesin removal occurs in prophase and is mediated by the phosphorylation of STAG2 (SA2) by polo like kinase 1 (PLK 1) and sororin by aurora kinase B (AURB) and cyclin dependent kinase 1 (Cdk1), however some of the cohesin remains attached at the centromere protected by Shugoshin like proteins (SGO1) and is finally removed before the onset of anaphase. SGO1 interacts with phosphatase PP2A and recruits PP2A to the centromere and this recruitment is believed to keep STAG2 in a dephosphorylated state preventing removal of cohesin from the centromere. WAPL and PDS5 are also essential for removal. It is thought that the phosphorylation of STAG2 causes changes in cohesin, which facilitates the dissociation of cohesin by WAPL and opening of the ring. Apart from the prophase pathway, the centromeric/pericentric cohesin in higher eukaryotes and bulk of the cohesin in yeast is removed by cleavage of RAD21/Scc1 by separase. Till the onset of anaphase, activity of separase is inhibited by its association with securin. In vertebrate cells separase is additionally inhibited by Cdk1 phosphorylation and its binding to Cdk1 subunit cyclin B. Activation of APC/C (Anaphase Promoting

Complex/cyclosome) by Cdc20 leads to ubiquitin mediated degradation of securin and cyclin B (Mehta et al., 2012) (Fig 1.9)

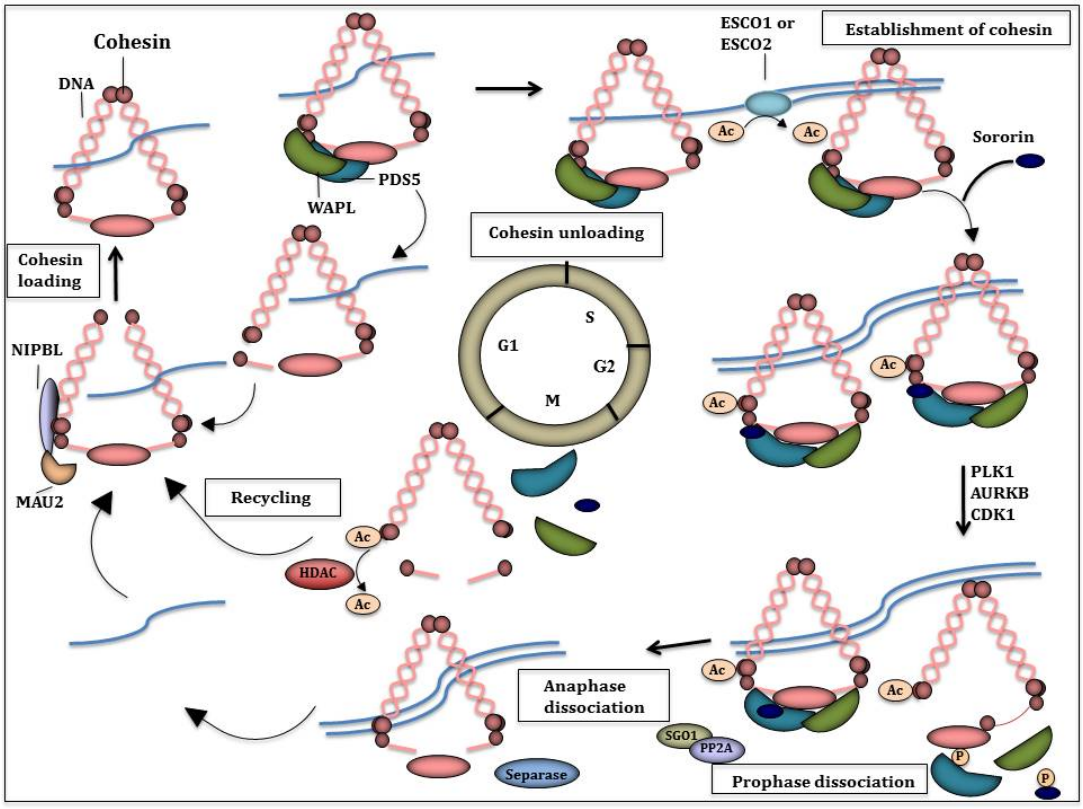


Figure 1.9: Cell Cycle regulation of Cohesin. Cohesin is loaded onto chromatin in early G_1 phase by the Nipped B like protein (NIBPL)-MAU2 heterodimer. Binding of PDS5 and Wings apart-like protein homologue (WAPL) to cohesin through RAD-21 and stromal antigen (SA) causes its unloading. DNA exits the complex through the interface created by SMC3 and amino terminal of RAD21. During DNA replication loading of cohesin on sister chromatids is promoted by acetylation of SMC3 by ESCO1 and ESCO2 acetylase (Ac) in the amino terminal of SMC3 at position K105 and K106. This acetylation causes recruitment of sororin which bounds to PDS5 and displaces WAPL although WAPL remains attached to complex. In prophase most of the cohesin dissociates from the sister chromatids which is caused by the phosphorylation of SA-subunit by polo-like kinase 1 (PLK1) and of sororin by aurora-kinase B (AURB) and cyclin dependent kinase 1 (CDK1) however some of the cohesin remains attached at the centromere and is prevented from dissociation by the concomitant action of shugoshin1 (SGO1) and protein phosphatase (PP2A). The remaining cohesin is removed at the beginning of the anaphase by the cleavage of RAD21 by separase. The cohesin complex released during mitosis can be recycled in G1 after deacetylation by histone deacetylase 8 (HDAC8) that removes acetyl group from SMC3.

1.5 Functions of Cohesin

1.5.1 DNA Repair/ DNA damage checkpoint regulation

The earliest evidence that cohesin is involved in DNA double strand break (DSB) repair came from the results in *S. pombe* where a cohesin mutant was rendered sensitive to γ radiation (Birkenbihl and Subramani, 1992). In fact the role of cohesin in DNA DSB repair was established before its involvement in sister chromatid cohesion. Scc1 depleted in the DT-40 chicken cell line, cell lines derived from breast cancer patients with impaired *RAD21* and depletion of *RAD21* from HeLa cells confirmed that cohesin is important for repair in higher eukaryotes.

It was first shown by Yokomoris group that cohesin is recruited to laser induced DNA damage sites in human cells in an MRE11, RAD50 dependent manner (Kim et al., 2002). The enrichment of cohesin at DNA DSB was also confirmed by Koshland and Sjogren groups in yeast and it was found that damage induced cohesion was not restricted at the DSB but spread throughout the genome in G_2 after genome duplication (Strom et al., 2007, Unal et al., 2007). This damage-induced cohesion (DI) was controlled by DNA damage response factors like Mec1, Tel1, Mre11, $_{\Upsilon}$H2AX and cohesin regulators (Scc2, Eco1, and Smc6) and was independent of replication. Further work done by Koshland and others found that in response to a DSB, Chk1 phosphorylates Scc1 at Serine (S83). This phosphorylation augments the acetylation by Eco1 at two positions, Lys84 and Lys210. The acetylation counteracts the effect of Wap1 in establishment of cohesion. Furthermore Smc3 acetylation by Eco1 was uniquely required for S phase cohesion and not for DI cohesion. So far evidence for DI cohesion has only been observed in yeast. Several lines of evidence show that DI cohesion might be a conserved mechanism in higher organisms. Sororin was found to be required for efficient DSB repair during G2 in sororin depleted HeLa cells and decreased inter-

sister chromatid distance was observed after DSB in DT40 chicken cells consistent with the establishment of DI cohesion. ChIP (chromatin immunoprecipitation) followed by deep sequencing (Chip-seq) revealed that ionising radiation (IR) triggered an Esco-1 dependent increase in acetylation of Smc3 and genome wide reinforcement of cohesin binding at pre-existing sites. (Kim et al., 2010, Wu and Yu, 2012).

Cohesin has also been found to be involved in intra S phase checkpoint in human cells. In response to DNA damage, ATM and ATR phosphorylated Smc1 at two serine residues S957 and S966 and this phosphorylation was found to be important for intra S phase checkpoint. Smc3 was also found to be phosphorylated at S1083 (Luo et al., 2008). How the phosphorylation of cohesin activates intra S phase checkpoint remains unclear. Cohesin has also been implicated in G2/M DNA damage checkpoint in human cells. Depletion of Scc1 by RNAi led to defective recruitment of 53BP1 to DNA damage foci and weaker Chk2 activation. Cohesin was also found to be required in Chk2 activation in G1 prior to replication (Watrin and Peters, 2009).

Cohesin's role in DSB repair is attributed to the fact that it can hold the sister chromatids together. This is important in terms of HR (Homologous Recombination) where the intact sister chromatid is held in close proximity with the broken DNA and can be used as template. A four- fold reduction of Scc1 or Smc3 decreased survival in response to radiation and also increased recombination between homologues leading to loss of heterozygosity (Covo et al., 2010). Cohesin may also regulate the choice of repair between HR and NHEJ pathway (Schar et al., 2004).

1.5.2 Gene Regulation

Cohesin regulates eukaryotic gene expression but the mechanism is poorly understood. The first evidence that cohesin is involved in gene regulation came from studies in Drosophila where the Scc2 ortholog *Nipped-B* was found to act as a modulator for the enhancer/promoter interaction at the *cut* and *Ubx* (Ultrabithorax gene) (Rollins et al., 1999). Evidence that cohesin has a gene regulatory role in vertebrates came from the work in zebra fish where the expression of *runx1* (Runt - related transcription factor 1) and *runx3* genes were abolished in early embryos mutant for *rad21*. Cohesin also regulates gene expression in post mitotic tissue. In Drosophila, a mutational screen found *SMC1* to be important for pruning of post mitotic neurons.

Cohesin associates with CTCF (CCCTC binding factor), which is an integral component of the c-myc insulator element MINE that separates transcriptionally active c-myc from the chromatin that bears features typical for heterochromatin. Several studies revealed that much of the cohesin co-localizes genome wide with CTCF. The fact that CTCF forms loops has already been confirmed by chromosome conformation capture (3C) assay. Dependence of CTCF based long range interaction on cohesin was first demonstrated for the mouse Interferon gamma locus (*lfng*) (Hadjur et al., 2009). In association with CTCF, cohesin mediates looping at a number of other genes including β globulin gene, X chromosome inactivation and *HoxA* locus in mice. At the β globulin locus, depletion of NIPBL or cohesin prevents the activation of globulin gene and formation of loops (Wang et al., 2011). *Wendt et al* showed that 89% of CTCF sites overlapped with Scc1, and Scc1 is required along with CTCF for genomic imprinting at the *Igf2/H19* locus (Wendt et al., 2008). In another study, *Rubio et al* using a proteomic approach showed that STAG1 (Scc3/SA1) subunit of cohesin interacts with CTCF at c-myc insulator

elements and CTCF recruits cohesin to the ICR (Imprinted Control Region) of the *Igf2/H19* locus in an allele specific manner.

Studies undertaken have uncovered two distinct cohesin binding sites that maybe involved in the gene regulation function of cohesin. Strong cohesin binding sites that coincide with CTCF such as the imprinting region where long term integrity and topology of the loops are important may require cohesin along with CTCF as a positioning partner. In contrast, numerous weak cohesin sites map to active promoters and enhancers where cohesin is colocalized with *NIPBL*(Merkenschlager and Odom, 2013). CTCF/cohesin binding sites (CBS) were found as mutational hotspot in a study to unravel mutational signature pattern at non -coding cancer genome. The findings provided clues to a number of predicted factors that contribute to CBS mutations. CBS mutations arise under abnormal conditions and tend to occur in late and special replicating regions such as the origin of replication. A preferential mutational signature T.A>C.G and T.A> G.C was observed at CBS sites which varied in gastrointestinal and mutated TP53 tumor types. Their work identified a new and unexpected class of cancer associated mutations that require vigorous effort to find its cause and consequences (Katainen et al., 2015).

Cohesin co-localises with estrogen receptor α in the MCF-7 cell line and with liver specific transcription factor HNF4A (Hepatocyte Nuclear Factor 4) in a hepatic cell line. Although the functional significance of cohesin binding in MCF-7 is not known, depletion of cohesin prevents estrogen responsive transition of breast cancer cells from G0/G1 indicating that cohesin influences the physiological estrogen response.

In mouse embryonic stem cells, cohesin and the mediator complex mediate long range interaction between enhancers and the transcription factors at the promoter

regions of pluripotent genes such as *Oct4* (Octamer binding transcription factor4), *Sox2* (Sex determining region Y– box 2) and *Nanog* that contribute to pluripotency and self-renewal by activating their own transcription and other genes involved in early development. Depletion of cohesin, mediator or Nipbl has same effect on embryonic stem cell as loss of *Oct4* suggesting that they are important for maintaining expression of key pluripotency transcription factors (Dorsett, 2011, Rhodes et al., 2011). However recent study has shown that cohesin depletion does not affect the expression of most of the pluripotent genes. Chromatin conformation capture (3C assays) revealed that the enhancer promoter interaction around pluripotent genes still remained strong even after 24hr of RAD21 depletion however induce DNA damage or DNA damage accumulation due to proliferation can affect gene expression (Gupta et al., 2016).

In concert with Polycomb group (PcG) silencing proteins cohesin restrains gene expression. Restraint genes coding for transcription factors like the *invected-engrailed* or *Enhancer of split* [E (spl)-C] are expressed at low to moderate levels in Drosophila. Such genes are bivalent having both the H3K27me3 (H3 lysine 27 trimethyl) methylated silent mark as well as H3K4me3 (H3 lysine 4 trimethyl) modification associated with transcriptionally active genes. Reduction of cohesin or polycomb strongly increases transcription of genes targeted simultaneously by cohesin and polycomb. Notably more than half of the 200 genes that increase in expression in mouse embryonic stem cells with cohesin knockdown are bivalent which maybe an important mechanism of gene regulation by cohesin in embryonic stem cell (Dorsett, 2011; Dorsett, 2012; Kagey, 2010).

In *Saccharomyces cerevisiae,* cohesin binds the heterochromatin region at the centromeric and telomeric regions and interacts with the protein that binds to these

regions. Cohesin is involved in the establishment and maintenance of the boundaries at HMR (Hidden mat right) silent mating locus. Mutations in SMC1 component of cohesin allow the SIR (Silent Information Regulator) silencing proteins to spread beyond the boundary causing silencing of the adjacent region (Dorsett and Strom, 2012, Rubio et al., 2008).

A study was undertaken to compare the distribution of cohesin in two different tissues (cortex and pancreas) in mouse and to study the effect on transcriptional regulation and chromatin architecture. Chromatin contacts were studied at protocadherin (Pcdh) and Regenerating islet-derived gene locus. Complete ablation of cohesin had a very reduced effect on chromatin contacts in the Pcdh clusters between the wildtype and SA1 null embryos in brain. In contrast to Pcdh the Reg locus seems to be sensitive to decrease in cohesin levels. Impaired homeostasis of the pancreas due to altered expression of the Reg genes made SA1 heterozygous mice more prone to pancreatic cancer (Cuadrado et al., 2015). In fission yeast Swi-6 ortholog of heterochromatin protein 1 (HP1) has been found to be responsible for association of cohesin with heterochromatin regions which leads one to speculate that expression of genes at this loci is affected by accumulation of cohesin at these loci.

1.5.3 Chromatin remodelling

Using a comprehensive protein interactome *Panigrahi et al* found cohesin interacts with chromatin remodelers like Chromodomain helicase DNA binding protein, (CHD4), the histone methyl transferase (SETD3), linker histone H1 and its chaperone Nucleosome assembly protein 1-like 1 (NAP1L1)(Panigrahi et al., 2012). Cohesin associates with chromatin remodeller SNF2h on chromosome arms in *vivo.*

These observations predict a role for SNF2h in mediating association of cohesin and chromatin (Hakimi et al., 2002).

1.5.4 Centrosome duplication

Cohesin is involved in faithful centrosome duplication. Evidence for this comes from studies on separase that was found to be responsible for centrosome duplication and centriole disengagement. Furthermore in HeLa cells depletion of RAD21 causes premature separation of centrioles. A splice variant of shugoshin (sSgo1) has been found to be required for protection of centrioles in mammalian cells. Work done by *Schockel et al* have shown that ectopic activation of separase or depletion of shugoshin (Sgo1) results in both premature sister chromatid separation and centriole disengagement in human cells (Schockel et al., 2011). Clarke and co-workers have also reported the importance of cohesin subunits (Rad21, Smc1 and Smc3) in bipolar mitosis suggesting that the cohesin ring is involved in maintaining the spindle pole integrity (Diaz-Martinez et al., 2010).

1.6 Cohesin and Meiosis

Meiosis- specific cohesin complexes consist of two structural maintenance of chromosomes (SMC1α/ SMC1β and SMC3) the α-kleisin subunit (Scc1/Rad21) which is replaced by Rec8 in yeast and metazoans, Rad21L in mammals and C(2)M in Drosophila and SA subunit STAG3. It has been observed that meiotic cohesin is involved in other cellular functions and not in sister chromatid cohesion.

1.6.1 Programmed Double strand break repair (PDSB)

At meiosis, cohesin is involved in programmed DSB repair required in the generation of chiasmata, a structure formed due to reciprocal exchange between non-sister chromatids. This differs from mitosis where repair is activated to maintain the genomic integrity against perturbation and involves recombination

between sister chromatids. The difference might have evolved due to the recruitment of a different kleisin subunit Rec8 in meiosis as compared to (Scc1/Rad21/Mcd1) in mitosis. Apart from its role in PDSB (Programmed Double strand break repair) Rec 8 might be involved in creating PDSBs as it has been found to colocalize with Spo11, the enzyme responsible for creating PDSB in yeast (Kugou et al., 2009). Also co-purification of Rec8 from different organisms with Rad51/Dmc1 which coats ssDNA required for strand invasion during PDSB repair also argues for involvement of Rec8 in events downstream of PDSB formation (Katis et al., 2010).

1.6.2 Pairing of Homologous Chromosomes

During meiosis I, pairing of homologous chromosomes takes place through the formation of the synaptonemal complex (SC). Cohesin proteins (Smc1, Smc3, STAG3) have been found to recruit recombination proteins such as Dmc1 (DNA meiotic Recombinase1) and Msh4 (Mut S protein homolog 4) on the chromosomes and can promote synapsis even in the absence of axial elements that are required for formation of the synaptonemal complex. Rad21L, a meiotic specific kleisin subunit is involved in initiating synapsis and recombination between homologs. Deficiency of Rad21L in male mice leads to failure in completing synapsis and eventually leads to azoospermia and infertility, whilst in female mice, absence of Rad21L leads to age related sterility (Herran et al., 2011). Even deficiency of Pds5 which is an accessory protein required for maintenance of cohesin has found to lead to failure in pairing and formation of synaptonemal complex, but no defects in sister chromatid cohesion. A recent study was conducted to assess genetic interactions between Stag3 and the α-kleisin subunits RAD21L and REC8 by constructing double knockout mouse model of *Stag3*, RAD21L and *Stag3* and REC8 and compare them

to REC8 and RAD21L double mutants. Stag3 Rad21L were found to be essential for axis formation between the chromosomes while Stag3 and REC8 were the main complex responsible for centromeric cohesion The ratio of centromere to chromatin signal was high for RAD21L single mutant suggesting that RAD21L is essential for chromocenter clustering unlike RAD21L, Rec8 double mutant (Ward et al., 2016).

1.6.3 Monoorientation of sister kinetochore during meiosis 1

One of the feature of meiosis 1 is that the sister kinetochore of each homolog are attached to the same spindle pole during metaphase 1(monoorientation) (Fig 1.10). It has been observed in *S.pombe*, *Oryza sativa* (rice) and *Arabidopsis* that during meiosis1 mutated/null Rec8 leads to biorientation of sister kinetochores as well as SCC defects (Chelysheva et al., 2005, Watanabe and Nurse, 1999). Interestingly in higher eukaryote (mice) Rad21 (mitotic homolog) is also expressed and plays a role in monoorientation. Consistent with this both Rad21 and shugoshin (Sgo2) have been involved in sister kinetochore association during meiosis 1 in mouse spermatocyte.

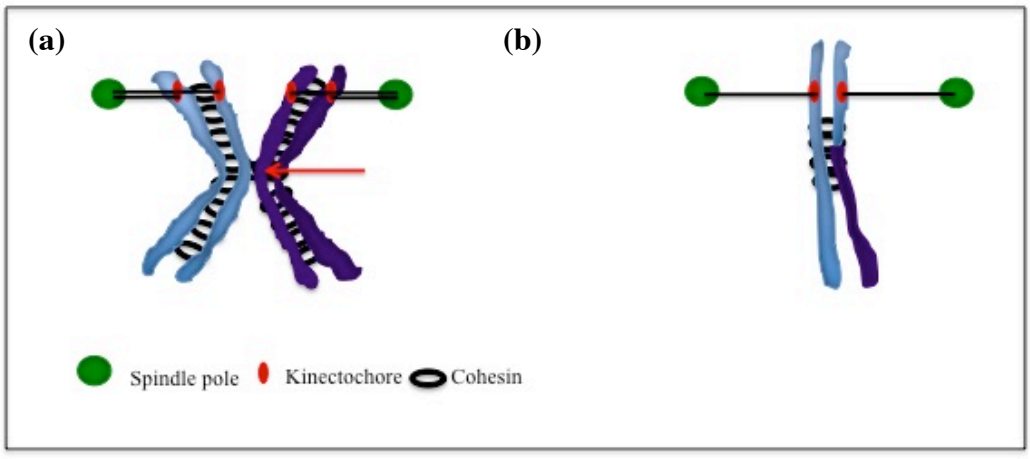


Figure 1.10: Kinetochore orientation by cohesin in meiosis. (a) A bivalent linked at the chiasmata (red arrow) by cohesin. Cohesin orients sister kinetochore towards the same pole of the spindle while homologous kinetochore are oriented towards the opposite pole in meiosis I (b) One of the two chromosomes derived from bivalent in meiosis II. Pool of cohesin around centromere helps in bipolar orientation of sister kinetochore in meiosis II.

1.7 Cohesinopathies

Cornelia de Lange Syndrome (CdLS) is the most common dominantly inherited congenital malformation disorder occurring in 1 in 10,000 live births. In 65% of the Cornelia de Lange probands it occurs due to a heterozygous mutation in the *NIPBL* gene located at 5q13 chromosome and the severe clinical features arise due to deletions or truncations in *NIPBL* (Dorsett and Krantz, 2009). As mutations in *NIPBL* account roughly for only more than half of CdLS, genetic screens was conducted to look for mutations in other genes with similar functions in a large cohort of CdLS cases without *NIPBL* mutations, which led to the identification of 5% of cases with a missense or small inframe deletion in *SMC1A* X linked gene, (Borck et al., 2007, Musio et al., 2006)<1% mutations in *SMC3* (Deardorff et al., 2007). Since then mutations in other cohesin genes like *RAD21* and *HDAC8* have been identified (Deardorff et al., 2012).

Roberts syndrome (RBS) and SC phecomelia are homozygous or compound heterozygous autosomal recessive disorder caused by a mutation in *ESCO2* that encodes a member of the acetyl transferase family (Schule et al., 2005). Mutations in *ESCO2* either cause complete loss of protein or loss of acetyl transferase activity. Mutations in *ESCO1* have not been reported, as they are lethal. SC phecomelia is a milder form of Roberts syndrome in terms of physical defects and mental retardation.

Although mutations in cohesin and cohesin associated genes cause these two syndromes the underlying etiology for these two syndromes can be different (Skibbens et al., 2013). Most CdLS patient cells do not show any defects in chromosomal segregation but rather exhibit gene dysregulation. Direct evidence to support this altered gene regulation comes from the study of transcriptional regulation of *HOXD* gene in mammals which is responsible for proximal-distal limb

patterning. Expression of *HOXD* gene, which is found to be reduced in CdLS patients, is found to be controlled by cohesin which brings the transcriptional control element located 200 kb away close to the transcriptional start site (Dorsett, 2009, Spitz et al., 2003). Gene dysregulation has also been observed due to mutation in cohesin loader in drosophila and zebra fish (Muto et al., 2011). Mice heterozygous for a mutation in Nipbl showed characteristics of CdLS. These phenotypes arose despite a decrease in Nipbl transcript levels of only approximately 30%, implying extreme sensitivity of development to small changes in Nipbl activity. Further gene expression profiling demonstrated that Nipbl deficiency leads to modest but significant transcriptional dysregulation of many genes including the protocadherin beta (Pcdhb) genes required for cell surface diversity generation in nervous system (Kawauchi et al., 2009). Consequently *Horsfield et al* has shown that mutation in Rad21 leads to developmental delay in zebrafish due to reduced expression of runx1 and run3 genes (Horsfield et al., 2007).

In a study to address how *ESCO2* brings about RBS, early embryonic lethality was observed in mice deficient in *ESCO2* in contrast to human patients with RBS. The reason behind this disparity has not been resolved. Deletion of both the copies of *ESCO2* in mice embryonic fibroblasts (MEF) led to chromosome missegregation and apoptosis. In a study conducted by *Wheelan et al ESCO2* was found to be important for cohesion at heterochromatin regions. In collating these results it has been proposed that developmental defects observed in RBS/ SC phecomelia may be due to cohesin dependent gene expression changes in heterochromatin regions (Whelan et al., 2012). In another study by *Bose et al* with budding yeast bearing mutations analogous to human cohesinopathies (eco-1W216G and smc1-Q843Δ) showed defects in ribosome biogenesis and protein translation but no defects in

chromosome segregation (Bose et al., 2012). Detailed analysis showed that both the mutants produce less ribosomal RNA that is expected to constrain ribosome biogenesis. Thus these results demonstrate that cohesin can regulate gene expression and cause disease ethology by altering general translational efficiency.

Recently a new cohesinopathy called Warsaw breakage syndrome has been identified which is caused by a defect in DDX gene (also known as *CHLR1*) a member of XPD helicase and could be a novel partner of cohesin pathway probably involved in DNA repair and holding the sister chromatids together (van der Lelij et al., 2010).

Enrichment of *SMC1A* was found at dysregulated genes in Cornelia de Lange Syndrome (CdLs) cell lines. Mutant cohesin impaired the occupancy of both Pol II and transition to the Pol II elongating form providing a molecular mechanism for the typical altered transcription profile observed in CdLs (Mannini et al 2015).

More recently *STAG2* was found to be a dosage sensitive gene, copy number gains of which was responsible for causing a novel cohesinopathy with Intellectual disability, behavioural problems and other clinical presentations. A large number of genes were dysregulated due to *STAG2* copy number gains one of which among them was oligophrenin1 (*OPHN1)* that is highly expressed in the brain and important for regulation of dendritic morphogenesis and synaptic plasticity. *OPHN1* is located on chromosome Xq12q13.1 and duplication encompassing *OPHN1* is associated with global developmental delay, autism (Kumar et al., 2015).

1.8 Cohesin as targets for therapeutic intervention

Overexpression of *SMC1A* has been found in triple negative breast cancer and supports the idea that in normal cells *SMC1A* biosynthesis is tightly regulated and that an imbalance in the amount of this protein may directly affect the cell survival

and other cellular functions. Inhibition of *SMC1A* showed more than three fold increased sensitivity in both BRCA wildtype and mutated triple negative breast cancer (TNBC) cell lines including basal-like and mesenchymal stem-like subtypes to PARP-inhibitor ABT-888 (Yadav et al., 2013). Further studies carried out showed loss of *SMC1A* associated with reduced proliferation and induction of apoptosis in colorectal and lung cancer cells. In particular associated loss of *SMC1A* mRNA and protein in colorectal cell line increased chemo sensitivity to oxiplatin in HT-29 cells (Li et al., 2016, Zhang et al., 2013).

Mutant high expression of RAD21 was found in tumorigenic cell line as compared to normal and immortalized breast cancer cell lines. Down regulation of RAD21 in breast cancer cell lines MCF-7 and T-47D had a significant effect on proliferation and viability, and further showed increased cytotoxicity to two DNA damaging agents etopside and bleomycin. The data from this study speculated RAD21 as a new target for breast cancer that could be used as an adjuvant to enhance the antitumor activity of traditional chemotherapeutic and radiation treatments (Atienza et al., 2005). A study conducted by *Gelot et al* demonstrated that cohesin complex prevents end joining of distal ends in S/G2 phase by suppressing both classic and alternate end joining C-NHEJ and A-EJ pathway. Further whole exome sequencing of RAD21 depleted SV40 fibroblast led to an increase in sequence insertion, duplication and deletion (Gelot et al., 2016). An investigation carried out by *Yun et al* found that reduction of cohesin decreased amplified copy number at genes *APIP*/*PDHX*/*CD44* in chromosomal unstable gastric cell lines and also of c-my existing at both HSR (homogenously stained regions) and DM (double minute) in chromosomal instable cell lines. Knockdown of *RAD21* decreased the recruitment of pre-replication complex at amplified genes and reduced the copy number of

amplified genes. Reduction in expression of cohesin also sensitized cells to DNA damaging agents like cisplatin and PARP inhibitors. In summary they predicted cohesin to be responsible for maintaining high copy number in cells with chromosomal instability (Yun et al., 2016).

Mutations in *STAG2* and *STAG3* (meiotic counterpart of SA proteins) have been found to confer resistance to B-Raf proto-oncogene, serine threonine kinase inhibitors (BRAFi) in melanoma. A study undertaken by *Shen et al* showed an increase in basal levels of phosphorylated (p)-ERK in these cells due to a significant decrease in ERK phosphatase called dual phosphatase specificity 6 (DUSP6) upon knockdown of *STAG2*. The promoter region of DUSP6 was found to contain a CTCF binding site which is responsible for controlling gene expression which was affected upon knockdown of *STAG2* however restoration of DUSP6 in the background of *STAG2* silenced shRNAs in melanoma cells enhanced the ability of BRAFi to inhibit ERK activities. Taken together their results strongly suggested that loss of *STAG2* inhibits CTCF mediated expression of DUSP6 leading to reactivation of MEK-ERK signalling in BRAFi-treated melanoma (Shen et al., 2016).

Inhibition of HDAC8 that is responsible for deacetylation of SMC3 by inhibitor PCI-34051 did not influence recruitment of estrogen receptor α at *SOX4* and *IL-2* genes but rather influenced cell cycle progression and survival. At higher doses of inhibitor the cell barely made into S phase and there was an alteration in the characteristic shape and cell viability was lost after 48hr of drug treatment (Dasgupta et al., 2016).

Genomic integrity is maintained by the co-ordinated activity of DNA damage response pathways, cell cycle checkpoints and alterations in transcriptional

regulation. The fate of genomic integrity can be threatened by deleterious double strand breaks that can lead to chromosomal rearrangements and apoptosis. Single agent PARP inhibitors have been exploited synthetically in various types of cancer especially in the area of breast and ovarian cancer however using PARP inhibitors in the context of cohesin mutations in myeloid malignancies needs to be explored.

1.9 DNA double strand repair Pathway

Of the many types of lesions that exist within a cell, DNA double-strand breaks (Bothmer et al.) are the most dangerous, continually threatening genomic integrity. DSBs are either created by endogenous sources that include, errors in DNA metabolism (e.g. replication across single- strand nicks and replication fork collapse), endogenous nucleases, programmed genome rearrangements and reactive oxygen species. Numerous DSBs of exogenous origin, both natural (e.g. cosmic rays, terrestrial background radiation, certain viruses) and man-made (e.g. weapons of mass destruction and diagnostic and therapeutic procedures) threaten the genome.

1.9.1 Challenges faced by Double-Strand-Break Repair Systems

Some of the requirements for an effective DSB repair include

a. Sensitivity: Rapid detection and repair of DSB before a catastrophic event i.e. cell death (apoptosis); unregulated cell division leading to tumor formation occurs.
b. Specificity: The repair system should be able to detect only DSBs but no nick, abasic sites; mismatches or interstrand crosslinks that require distinct repair systems.
c. High fidelity: The ability to repair DSBs without causing any collateral damage.

d. The ability to repair a variety of DNA ends including those produced by certain forms of irradiation and reactive oxygen species that are not directly ligatable.
e. The ability to coordinate the timings of repair with the physiological state of cell.

Typically cells uses two main mechanisms for DSBs repair homologous recombination (HR) which occurs in S phase and classical non-homologous end joining (C-NHEJ) which is dominant in G1 and sub S phases. Various alternate error prone pathways like alternate end joining (alt EJ) and single strand annealing (SSA) exist as well but contribute towards genomic rearrangement and oncogenic transformation. Members of the Phosphotidylinositol-3kinase-related kinases (PIKK) through their ability to phosphorylate a large number of substrates play important role in different stages of DSB signalling. These kinases are Ataxia-Telangiectasia-Mutated (ATM) and DNA-dependent Protein Kinase catalytic subunit (DNA-PKcs) critical for signalling of DSBs while Ataxia- Telangiectasia and Rad3 Related (ATR) is involved in response to DNA single strand breaks (SSBs) and stalled replication forks (Bartek and Lukas 2007), (Jazayeri et al., 2006). In mammalian system MRE11/RAD50/NBS1 (MRN) and Ku 70/80 sensors of DSBs that also aid in processing of DNA ends recognise DSBs and recruit ATM and DNA PKcs to site of DSBs (Carson et al., 2003, Difilippantonio et al., 2005, Mordes and Cortez, 2008). Homologous recombination (HR) repair is tightly coordinated with cell cycle progression, which is in large part governed by cyclin-dependent kinases (CDKs). CDK-mediated phosphorylation of C-terminal interacting protein (CtIP) and Nijmegen Breakage Syndrome 1 (NBS1) appears to be essential for MRN-mediated DNA-end resection (Limbo et al., 2007, Sartori et

al., 2007). MRN complex interacts with the N-terminal of ATM and recruits it to DSBs. It is also required for activation of ATM. ATM exists as inactive dimers or multimers in undamaged cells however auto phosphorylation in trans leads to its dissociation into active monomers. Serine(S) S1981 was the first auto phosphorylation site to be identified (Bakkenist and Kastan, 2003). Additional phosphorylation sites have been documented at S367 and S1893 that may contribute to activation process. Active ATM then promotes recruitment of C-terminal interacting protein (CtIP) to damaged sites where it interacts with and stimulates the nuclease activity of nuclease meiotic recombination 11 (MRE11) to initiate strand resection and generate short tracts of ssDNA. ATM substrates are rapidly phosphorylated after activation of ATM and include structural maintenance of chromosome-1 (SMC1), (Nijmegen breakage syndrome 1) NBS1, checkpoint kinase 2 (CHK2), p53BP1, breast cancer early onset-1 (BRCA1) mediator of DNA damage checkpoint protein-1 (MDC1) and γ-H2AX (Lavin, 2008). In response to DNA damage ATM, ATR and DNA-PKcs phosphorylates H2AX a histone H2A variant on its serine 139 (γ-H2AX) (Rogakou et al., 1998) MDC1 recognizes γ-H2AX and binds to it through its BRCT domain which further promotes accumulation and retention of active ATM and MRN complexes to γ-H2AX containing chromatin surrounding the sites of DNA damage (Stewart et al., 2003). In undamaged cells MDC1 exists in a complex with MRN however following ATM activation MDC1 and its interacting MRN complex are recruited to γ-H2AX. End resection occurs in two phases: an initial phase called end clipping and a second phase called extensive resection. The initial phase involves MRE11 and CtIP which removes a limited number of bases (20 bp in mammalian cells and 100-300 bp in yeast) Reviewed in (Ceccaldi et al., 2016). The second phase involves extensive end

resection by helicase and endonuclease (BLM, CtIP, EXO1, DNA2, WRN) generating 3' overhangs (Sturzenegger et al., Symington and Gautier, 2011). Upon DNA damage the eukaryotic ssDNA binding protein replication protein A (RPA) initially competes with RAD51 for single strand binding, however RPA also has a pro recombinogenic role once RAD51 is loaded on ssDNA it favours presynaptic formation by eliminating secondary structures and by protecting DNA ends from degradation (Chen et al., 2013). RAD51 paralogs BRCA2 and RAD52 favour RPA displacement, RAD51 nucleofilament formation and strand exchange activity (Fig 1.11 b). BRCA2 is the main mediator of RAD51 nucleofilament formation and strand exchange that it does through a series of eight evolutionary conserved motifs called BRC repeats. BRCA2 BRC domains promotes ssDNA binding of RAD51 by disrupting self assembled RAD51 oligomers and favouring one to one binding of RAD51 monomers. BRCA2 binding also stabilizes ATPase activity of RAD51 thereby supporting the ssDNA binding activity of RAD51. In a second step the BRCA2 C terminal domain binds to RAD51 oligomers in the context of the RAD51-ssDNA helix and thus promoting nucleofilament growth and participating in strand invasion. This gives rise to D-loop intermediates where the 3'end of invading strand primes DNA synthesis using the homologous chromosome as a template. The invading strand is disengaged after DNA synthesis and annealed with the second end leading to localized conversion without crossover in mitosis.

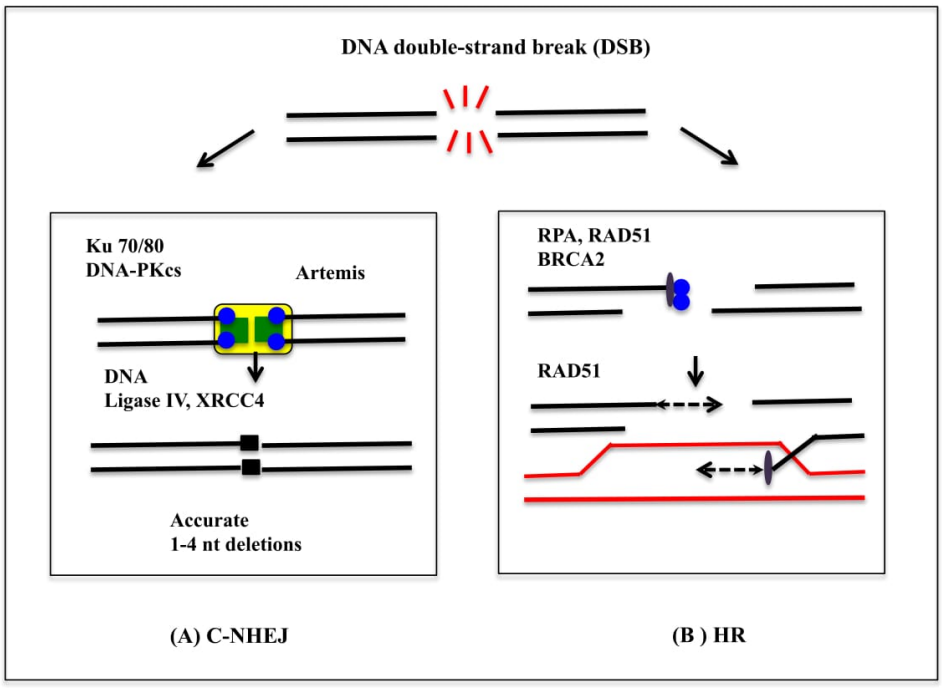


Figure 1.11: Pathways to repair DNA Double Strand Breaks (DSBs). Selection of the pathway to repair DSBs primarily depends on end resection. If end resection is blocked then repair of DSBs occurs through NHEJ pathway (A) that involves binding of Ku70/80 and DNA-PKcs to broken ends. DNA-PKcs phosphorylate a large number of substrates like Artemis, XRCC4, DNA Ligase IV, followed synapsis of broken ends and ligation followed by end processing and DNA ligation. Once resection has occurred repair takes place through HR (B) that involves binding of RAD51 (recombinase) to resected ends followed by strand invasion using the intact strand as template. After DNA synthesis has occurred the invading strand is disengaged followed by ligation with the broken ends and results in localized conversion without crossovers.

Classical NHEJ is initiated by Ku70/80 end binding which protects broken ends and causes the recruitment of catalytic subunit of DNA-dependent protein kinase (DNA-PKcs) (Fig 1.11 a) (Jacobs et al., 2011). Ku initially interacts with the distal termini of DSB and protects them from end resection (Downs and Jackson, 2004). Recruitment of DNA-PKcs translocates Ku inwards by about one helical turn allowing interaction of DNA-PKcs 10bp upstream at both termini (Yoo and Dynan, 1999). DNA-PKcs upon binding causes phosphorylation of a large number of substrates including Ku, Artemis, XRCC4, Ligase IV and XLF (XRCC4 like factor) followed by synapsis of DNA ends and recruitment of end processing and ligation enzymes (Neal and Meek, 2011). Phosphorylation of DNA-PKcs either trans-auto

phosphorylation (Chan et al., 2002) or phosphorylation by ATM promotes its dissociation from DNA end. DNA ends in classical NHEJ that have a 5'phosphate and a 3'OH group can be ligated directly, however complex ends produced by reactive oxygen species or irradiation cannot be ligated directly and hence additional enzymes are required to process these termini for ligation. Artemis, a nuclease processes free ends for more efficient ligation during cNHEJ (Moshous et al., 2003). Once the endonuclease activity of Artemis is activated by DNA-PK, the complex is required to remove single stranded DNA overhangs containing damaged nucleotides.XRCC4 has been shown to interact with polynucleotide kinase/phosphatase (PNKP) that phosphorylates 5' OH group and dephosphorylates 3'-phosphate termini providing the correct end groups for DNA ligation. Extension of the DNA ends requires a particular DNA polymerase that fills in the end at or near the site of DNA damage. Four classes of DNA polymerase have been recognised in eukaryotes A, B, X and Y families (Burgers et al., 2001). Three members of the pol X family have been associated with mammalian NHEJ. pol λ, pol μ and terminal deoxynucleotidyl transferase (TdT) which all share the BRCT (BRCA1 C-terminal) domains essential for complex formation between the polX members and core cNHEJ factors at DNA ends (Ramsden, 2011). The final step of joining the DNA ends in cNHEJ pathway is carried out by DNA XRCC4, Ligase IV and XLF complex. Ligase IV has a N-terminal catalytic domain and interacts with the α helix of XRCC4 via a region between the two C-terminal BRCT domains. Binding of XRCC4 stabilizes and stimulates the activity of Ligase IV (Wu et al., 2011).

1.9.2 Synthetic lethality

The term Synthetic Lethality was coined by Calvin Bridges in 1922. The concept arises from the genetic studies done in Drosophila where loss of one gene function is tolerated by overreliance on another gene in a redundant pathway. Synthetic lethality occurs when the gene from the redundant pathway is also deleted or its product inhibited. Although conventional chemotherapy has been used to treat cancer its range and effectiveness has largely been restricted due to its side effects. The growing capability of cancer cells to resist chemotherapeutic drugs is one of the major drawbacks besides inability to distinguish between normal and malignant cells. Also overdosage of drug and immunosuppression has harmful side effects. The use of targeted therapy helps to overcome the drawbacks of conventional chemotherapy. Targeted therapy stops growth of cancer cells by interfering with specific target molecules responsible for carcinogenesis and tumor growth. The use of PARP inhibitors especially in the field of ovarian and breast cancer has revolutionized the field of targeted therapy.

1.10 PARP

Poly (ADP-ribose) polymerases (PARPs) also known as ADP-ribosyltransferases (ART) are members of a small family of proteins that play important roles in biology (Bock and Chang, 2016). Humans express 17 of these enzymes that catalyse the post translational modification of target proteins, a process termed PARylation (Luo and Kraus, 2012, Rouleau et al., 2010). PARP1 is the most abundantly present isoform and plays crucial role in DNA repair, epigenetic modulation of chromatin, regulation of genomic stability, modulation of cellular energy pools, the regulation of transcription and a distinct form of cell death called parthanatos (De Lorenzo et al., 2013).

1.10.1 Structure of PARPs

Members of the PARP family are characterized by the presence of a conserved 50 amino acid sequence found in the C-terminal domain that is highly conserved in eukaryotes and is also known as the PARP signature (Virag and Szabo, 2002). Outside of this domain, PARP enzymes are quite diverse and based on the presence of well characterized protein domains are grouped into subfamilies (Fig 1.12). PARPs utilize these domains to target them to specific cellular locations, to identify target proteins and to regulate the enzymatic activity of catalytic domain.

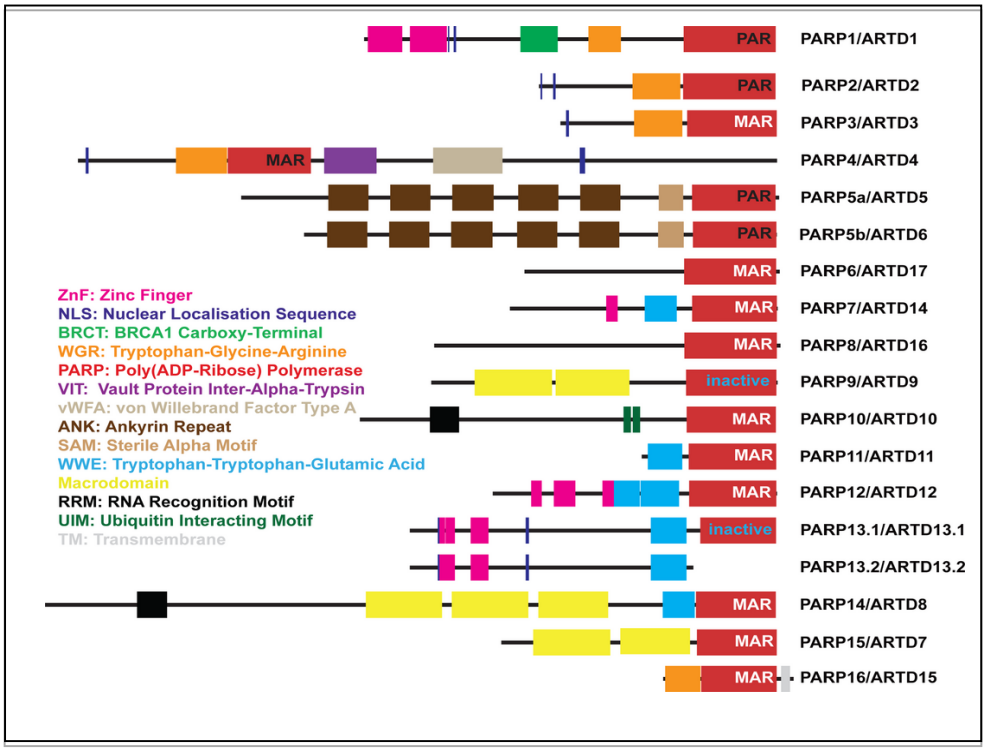


Figure 1.12: The PARP family. The 17 Human PARPs. PARP1 the mostly widely studied PARP enzyme has an N terminal DNA binding domain that contains zinc finger binding motif, the central automodification domain, with a BRCT motif and a C-terminal catalytic PARP domain that is present in most of the PARP enzymes. PARP 13.2 does not contain the PARP domain. (Bock and Chang 2016)

PARP1, a highly abundant nuclear enzyme discovered 50 years ago activated by DNA damage, plays an essential role in single strand base repair (SSBR) (Chambon et al., 1963). PARP2 has structural similarity and overlapping function with

PARP1.The catalytic activity of PARP1 increases by 10-500 fold after it is recruited to DNA damage sites, resulting in the synthesis of protein conjugated long ADP-ribose polymers. PARP enzymes use nicotinamide adenine dinucleotide (NAD^+) as a cofactor and transfer the ADP-ribose moiety to form mono or polyADP-ribose chain (pADPr) on proteins like histones, topoisomerase1, and DNA protein kinases. However, the bulk of pADPr is attached to PARP1 itself resulting in the recruitment of over a hundred other proteins, one of which is XRCC1 (X-ray cross-complementation group 1) that acts as a scaffold for DNA polymerases and ligase components of base excision repair machinery (BER) for both short and long patch repair (El-Khamisy et al., 2003). Formation of pADPr reduces the affinity of PARP1 and histones for DNA providing a mechanism for removing PARP1 from the damaged site and local modulation of chromatin compaction. Further polymer synthesis is also antagonized by two enzymes that hydrolyse pADPr, poly (ADP-ribose glycohydrolase) (PARG) and possibly ADP-ribose hydrolase 3 (ARH3) (Rouleau et al., 2010).

Apart from SSBR, PARP1 is also involved in several other biological processes. Rapid recruitment of Ataxia telangiectasia-mutated (ATM) to sites of DNA DSBs is dependent on pADPr synthesis suggesting that PARP1 acts as a facilitator of homologous recombination DSB repair. *Haince et al* envisaged a model were parylation of modified proteins or PARP1 itself facilitates the phosphorylation of ATM downstream targets i.e. p53, SMC1A and γH2AX (Haince et al., 2008). PARP1 competes with Ku in preventing high fidelity repair and promote mutagenic repair by alternate non-homologous end joining (Alt- NHEJ). DSB repair of the AID- induced genetic lesions generated during immunoglobulin gene conversion are repaired by Alt-NHEJ (Paddock et al., 2011).

PARP1 is involved in chromatin modulation and regulating the composition of chromatin. Binding of PARP1 to nucleosomes in the absence of NAD^+ leads to compaction of nucleosomal arrays into high order structures that are refractory to *invitro* transcription. Presence of saturating amounts of NAD^+ leads to automodification of PARP-1, de-compaction of chromatin and restoration of transcription (Kraus, 2008). PARP1 plays a role in altering chromatin composition by excluding Histone H1 from the promoters of some PARP-1-regulated genes possibly by competing with it to bind to nucleosomes.

PARP1 acts as a promoter specific co-regulator (either a coactivator or corepressor) for a number of sequence specific DNA binding transcriptional regulators such as Nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB) nuclear receptors, Oct-1, specificity protein 1 (Sp1) and others. In most cases, the DNA binding motif of PARP is responsible for recruitment of PARP-1 to relevant target promoters. PARP-1 can also act as a promoter specific exchange factor that promotes the release of inhibitory transcription factors and the recruitment of stimulatory factors during signal-regulated transcriptional response (Kraus, 2008). Parp1-/- showed hypersensitivity to DNA damaging agents particularly alkylating and ionising agents (Shall and de Murcia, 2000). Gene expression profiling in Parp1-/- mouse cells and breast cancer cell lines treated with PARP1 short-hairpin RNAs (shRNAs) reveal that PARP1 loss or down regulation alters the expression of many genes involved in cell cycle control and stress response such as p53 (Frizzell et al., 2009, Simbulan-Rosenthal et al., 2000). Among the upregulated genes from Parp1-/- fibroblasts were genes involved in the extracellular matrix or cytoskeletal proteins that have been reported to play roles in cancer initiation and progression, whilst being associated with normal or premature aging. Collectively these

observations implicate PARP1 in transcription as well as multiple aspects of DNA damage response.

1.10.2 PARP Inhibitors

The development of competitive inhibitors of PARP catalytic activity has become an area of active research and much recent excitement in PARP field. The focus is to develop specific, potent, effective and safe PARP inhibitors that may be used as research tools and for clinical therapies. The idea of synthetic lethality to target cancer was first highlighted by Hartwell (Hartwell et al., 1997) and colleagues and subsequently championed by Kaelin (Kaelin, 2005).The most pertinent example of synthetic lethality to treat cancers comes from tumor specific loss of tumor suppressor genes *BRCA1* and *BRCA2* in ovarian and breast cancer (Lord et al., 2015). *BRCA1* and *BRCA2* were originally identified as familial breast and ovary cancer predisposition genes. PARP1 is involved in the repair of SSBs during BER and inhibition of PARP1 leads to the accumulation of SSBs that are converted to DSB after replication or stall replication forks. Stalled replication forks and DSB, are generally repaired by DNA recombinase RAD51 dependent homologous recombination that requires BRCA1 and BRCA2. In the absence of BRCA1 and BRCA2, DSB and replication forks cannot be restarted and collapses causing persistent chromatid breaks. The repair of these breaks by alternate error prone pathways results in the accumulation of large number of chromatin breaks and aberrations leading to loss of viability (Fig 1.13).

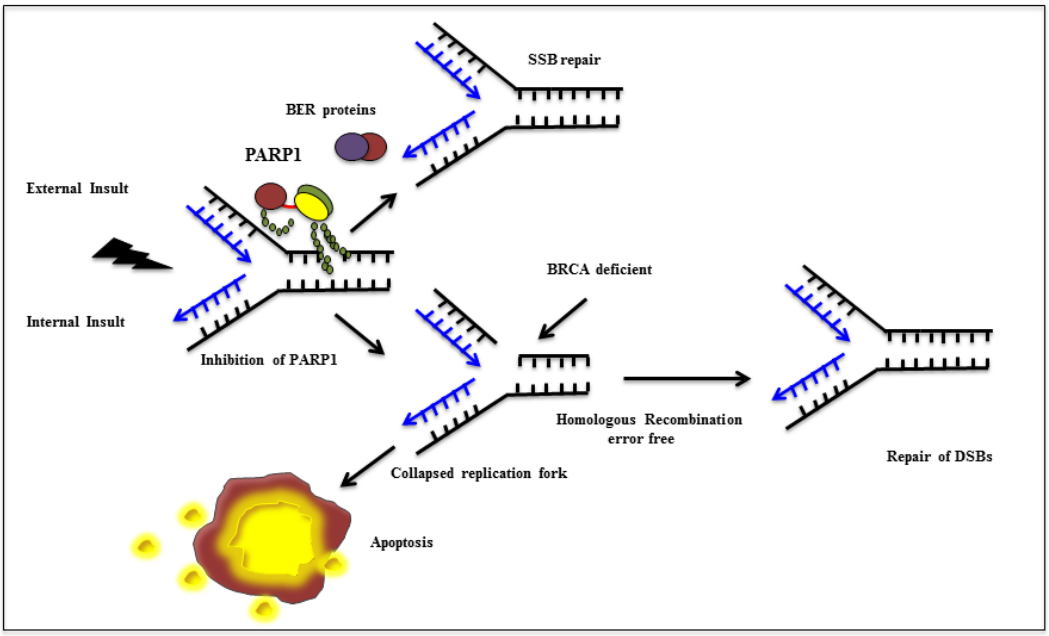


Figure 1.13: Concept of Synthetic Lethality according to *BRCA* status. External or Internal insults to genomic integrity results in the formation of SSBs. XRCC1 (X-ray repair cross complementing protein1) which forms a multimolecular complex with PARP1 is involved in base excision repair (BER) of SSBs. When exposed to PARP inhibitors SSBs accumulate which convert to DSBs and subsequent collapse of replication fork. In cells with intact BRCA these breaks are repaired by homologous recombination repair pathway leading to DNA repair and cell viability. Cancer cells with dysfunctional BRCA are deficient in HRR pathway and when exposed to PARP inhibitors DSBs accumulate leading to cell death.

In 2003, two groups demonstrated the potential of using PARP1 inhibition as a targeted synthetic lethal approach to treat BRCA –mutant cancers on the basis of this model (Bryant et al., 2005, Farmer et al., 2005).

3-aminobenzamide (3-AB) was the first PARP inhibitor (PI) to be developed, but it lacked the requisite potency and selectivity to be used as a research tool or in the clinic. All PARP inhibitors have an amide pharmacore and are catalytic inhibitors competitive to NAD^{+}. PARP inhibitors do not prevent binding of PARP1 to DNA but inhibit synthesis of pADRr polymer and DNA repair. They may further hamper repair as first suggested by Satoh and Lindahl by preventing the dissociation of PARP1 and thus physically preventing access of repair proteins. However, another mechanism of PARP-1 inhibition involving the trapping of DNA/PARP-1

complexes has been proposed where PARP-1 inhibitors (olaparib, velaparib, talazoparib) binds to the catalytic domain of PARP-1 and causes an allosteric conformational change in PARP-1, stabilizing the reversible association of PARP1 with DNA which results in persistence complexes at SSBs and conversion of SSBs into DSBs in HR deficient cells causing cell death (Murai et al., 2012, Murai et al., 2014).

1.10.3 PARP as chemo and radiosensitizers

PARP inhibitors (PARPi) have also been found to increase the cytotoxicity of DNA methylating agents, topoisomerase poisons and ionising radiations (IR) in combination. Several studies have investigated chemosensitization by second and third generation PARP inhibitors. In one of these studies, PARP inhibitor NU1025 was found to potentiate DNA strand breakage and cytotoxicity induced by topoisomerase I inhibitor camptothecin. Lack of PARP activation prevented enhancement of etopside mediated cytotoxicity or DNA strand breakage by PARP inhibitor NU1025. This differential effect between topoisomerase I and II inhibitors could either be cell specific or due to the differences in the nature of DNA strands breaks formed by the two topoisomerase inhibitors The results from this investigation paved way for the use of PARP inhibitors in combination with topoisomerase I inhibitor in anti cancer therapy(Bowman et al., 2001).

In another study, the PARPi, PD128763 and NU1025 were found to potentiate the DNA strand breakage and cytotoxicity induced by temozolomide, Both the inhibitors were found to be 40 and 50 fold more potent than 3 aminobenzamide (3AB) and benzamide (BZ) and could be used in micro rather than millimolar concentration (Boulton et al., 1995).

PARP is responsible for the repair of various types of DNA damage caused by radiotherapy. Radiosensitization by PARPi may be caused by the accumulation of SSBs that are converted to DSBs during replication as is supported by increased sensitivity of S-phase cells to IR by PARP inhibitor 4-amino-1, 8-napthalamide reviewed in (Curtin, 2014).

1.10.4 PARP Trials

In 2005, the first phase I clinical trial began that established preliminary clinical evidence of efficacy of using PARP inhibitors to treat BRCA mutant cancers. In an initial accelerated dose escalation phase, the maximum dose of PARP inhibitor olaparib (AstraZeneca/KuDOs) was well tolerated in 12/19 patients with BRCA1/2 mutated breast, ovarian and prostrate cancer with an objective response rate (ORR) of 47% and disease control rate (DCR) of 63%. Further myelosuppression and nervous system side effects were much milder than those elicited by standard chemotherapy (Kilburn and Group, 2008) A retrospective analysis showed that platinum sensitive patients showed the most favourable response to olaparib with a response rate of 69%.

On the basis of the promising results of phase I trials, two phase II trials in chemo-refractory breast and ovarian cancer used olaparib at two doses of 400mg and 100mg twice a day. The ovarian cancer cohort showed an effective response rate of 33% while a response rate of 41% [median progression free survival (PFS) of 5.7 months] was observed in breast cancer cohort with 400-mg twice a day (Audeh et al., 2010, Tutt et al., 2010). Progression-free survival (PFS) is defined as the time from random assignment in a clinical trial to disease progression or death from any cause.

Two additional phase II studies by *Germon et al* (Gelmon et al., 2011) and *Ledermann et al* (Ledermann et al., 2012) (commonly referred to as study 19) in sporadic advanced ovarian and triple negative breast cancers (TNBC) were initiated. In the Gelmon study, the presence or absence of BRCA mutation was used as a selection criteria and patients received 400mg of olaparib twice a day, sustained antitumor response which correlated with prior platinum sensitivity was seen even in the absence of BRCA mutation in the ovarian cohort while in the breast cancer cohort despite reduction in tumor and higher frequency of disease stabilization no sustained response was achieved either in BRCA mutated or non mutated group. The Ledermann study assessed the utility of olaparib as a maintenance therapy in high grade serous ovarian cancer (HGSOC) after showing an initial response to platinum. The PFS for germline or somatically occurring BRCA1 and BRCA2 was 11.2 months versus 4.3 months for patients who received placebo.

The approval of Olaparib in 2014 for the treatment of BRCA deficient ovarian cancers validated PARP-1 as an anticancer target and established its clinical importance in cancer therapy. Around nine PARP inhibitors are currently active in various clinical trials which show potent activity (IC_{50} in the nM range) but nonselective inhibition of PARP-1 and PARP-2 (Wang et al., 2016).

Varied PARP activity was also found in a cohort of 109 Chronic Lymphocytic Leukaemia (CLL) patients. Parp activity correlated with PARP1 protein expression, however no association was found between PARP activity and p53 status or ATM loss, Binet stage, IGHV mutational status or survival but significant correlation was found between Bcl-2 (antiapoptotic factor) and Rel A (an NF-κB subunit). Further culturing of CLL cells on an irradiated monolayer showed sensitivity to potent

PARP inhibitor talazoparib at nM concentrations independent of p53, ATM or HRR status and hence suggested PARP1 as a therapeutic target in CLL by PARP inhibitors (Herriott et al., 2015).

A substantial amount of work in the field of PARP has been done by *Gaymes et al* from our group. Initial work carried out by him proposed that chromosomal instability disorders like Fanconi anaemia and Blooms syndrome which show increase propensity to transform to leukaemia could be targets for PARPi therapy due to DNA repair defects. The data also suggested that PARPi can target cells with defects in DNA repair and signalling molecules rather than sole defects in homologous recombination. The chromosomal unstable cell lines showed sensitivity to PARPi and increase in non homologous end joining activity and proposed that increase non homologous end joining and loss of HR competency make chromosomal instability syndromes sensitive to PARP inhibitors (Gaymes et al., 2008). Subsequent study conducted by him showed that PARP inhibitors cause cell cycle arrest and increase apoptosis in primary leukaemic samples and leukaemic cell lines. The effect of PARP inhibitors were potentiated by the use of histone deacetylase inhibitors (HDAC) and supported the feasibility of phase I clinical trial alone or in combination in AML and high risk MDS (Gaymes et al., 2009). Further another study showed that microsatellite instability (MSI) is caused due to defects in the mismatch repair genes (MMR) and MSI dependent mutations in DNA repair genes like MRE 11 and CtIp confer sensitivity to PARP inhibitors (Gaymes et al., 2013).

Lastly work carried out by another group in our department showed that recessive transcriptional factors in AML like AML ETO and PML RARA showed strong sensitivity to PARP inhibitors as compared to MLL-fusions which were insensitive

to PARP inhibitors and found Hox-A9 responsible for the differential effect. In activation of Hox9 in AML1 ETO and PML RARA made them insensitive towards PARP inhibitor olaparib. Combined use of PARPi along with Hox-A9 inhibitor achieved selective killing of otherwise PARPi resistant MLL leukemic cell line revealing a novel venue to overcome Parp resistance in AML (Esposito et al., 2015)

A large number of inhibitors have been developed against common mutations in AML. 5-azacytidine and 5-aza2deoxycytidine (decitabine) are hypomethylating agents which have been found to reverse aberrant DNA hypermethylation leading to restored expression of critical tumour-suppressor genes and have been used against *DNMT3* mutations (Challen et al., 2011). Synthetic lethality has also been reported between IDHI/2 mutations and inhibitors of BCL in AML and are being used in clinical trials

Lastly work done by *Gaymes et al* in our group has shown that 15-20% of leukaemic patient sample cells in his cohort were sensitive to PARP inhibitors.

Identification of cohesin mutations in myeloid malignancies as target for therapeutic intervention will help in the selection of suitable candidates for PARP inhibitor treatment so a part of my work will be focussed to unravel this possibility.

Hence the aim of this study is

To determine the prevalence of occurance of cohesin mutations in a cohort of MDS patients and co-relate this with other genetic mutations, survival, response to therapy.

To determine the functional consequences of knockdown of *STAG2* in a myeloid cell line. Exploiting *STAG2* as a target for therapeutic intervention through identification of possible synthetic lethality between *STAG2* and PARP inhibitors.

Chapter Two

MATERIAL AND METHODS

Reagents

2.1.1 Antibodies

Primary Antibody	**Dilution used for western blotting**	**Manufacturer**
Rabbit anti-SMC1A	1:500	ab21583 abcam
Rabbit anti-SMC3	1:1000	ab9263 abcam
Mouse anti-RAD21		05-908 Millipore
Mouse anti-STAG2 (J-12)	1:500	sc-81852 Santa Cruz
Mouse anti-STAG1 (LL-16)	1:200	sc-81851 Santa Cruz
Rabbit anti-RAD51 (H-92)	1:50	sc-8349 Santa Cruz
Mouse anti-Phospho-γH2AX	1:50	JBW 301 Millipore
β-Actin	1:4000	ab8580 abcam
γ-Tubulin	1:3000	sc-7396 Santa Cruz
Secondary Antibody	**Dilution**	**Manufacturer**
HRP-conjugated anti-mouse antibody	1:3000	Sigma
HRP-conjugated anti-rabbit antibody	1:3000	Sigma
Anti mouse IgG FITC	1:200	Sigma
Anti rabbit IgG TRITC	1:200	Sigma

2.1.2 PCR, Gel Electrophoresis and Sequencing Materials

Agarose, molecular biology grade /electrophoresis grade	Sigma
BigDye Terminator v3.1 cycle sequencing kit	Applied Biosystems
BigDye XTerminator purification kit	Applied Biosystems
Ethidium Bromide	Sigma
Filtered and unfiltered tips	Star labs
Gel cleaning kit QIAEX II, gel extraction kit	Qiagen
GoTaq ®colourless Master Mix	Promega
Nuclease free Water	Sigma
O'Gene Ruler ladder mix (0.1 µg/µl)	Fermentas
6x Orange DNA loading dye	Fermentas
50XTris Acetate EDTA buffer (TAE buffer)	Sigma

2.1.3 DNA and RNA extraction reagents

Chloroform	Sigma
DNeasy blood and tissue kit	Qiagen
Ethanol	Sigma
2-Propanol (Iso-propanol)	Sigma
RNase free water	Qiagen
TRizol	Invitrogen

2.1.4 Cloning Reagents

BactoAgar	Becton Dickson
Bacto Yeast Extract	Becton Dickson
Carbenicillin solution 100mg/ml in 50% ethanol (10 ml)	Bioline
GenElute ™ HP Plasmid Maxiprep Kit	Sigma
One Shot ® Mach1™ T1R chemically competent *E.coli*	Invitrogen
Sodium Chloride (NaCl)	Sigma
Tryptone enzymatic digest	Sigma Aldrich
Wizard ® Plus SV Miniprep DNA purification kit	Promega

2.1.5 Tissue culture reagents

Dimethyl sulfoxide (DMSO)	Sigma
Eagles Minimum Essential Medium (EMEM)	ATCC
70% Ethanol (Fixing cells)	Haymankimia
Fetal Bovine serum	Sigma
Neubauer Improved Haemocytometer	VWR International
Penicillin/Streptomycin solution	PAA
Phosphate Buffered Saline (PBS)	PAA
Roswell Park Memorial Institute Medium (RPMI- 1640)	Thermo Fisher Scientific
Trypan Blue 0.4%, 0.85% NaCl	Lonza
TrypLE ™ Express Enzyme	Thermo Fisher Scientific

2.1.6 Transfection reagents

Amaxa cell line nucleofector kit V	Lonza
Ingenio cuvettes	Cambridge Bioscience
Ingenio electroporation solution	Mirus
Ingenio pipettes	Cambridge Bioscience
Polybrene	Sigma
Puromycin	Sigma

2.1.7 Plastic and glass ware

15 and 50 ml falcon tubes	VWR International Limited
25, 75 and 175cm^2 ventedtissue culture flasks	VWR International Limited
5ml, 15ml, 25ml serological pipettes	VWR International Limited
6, 12, 24, 96 well plates	VWR International Limited
10µl, 200µl and 1ml filter tips	Star labs
Coverslips	BDH coverslips
Cryovials	Greiner bio-one
1.5ml microcentrifuge eppendorf tubes	Star labs
Poly lysine coated slides	Sigma
Syringe filters 0.22 and 0.45µm	Millex
Triple layered flasks	Millipore

2.1.8 Plasmids

Expression plasmid	Plasmid description	Manufacturer
GIPZ STAG2 shRNA	Generation of lentivirus	Thermo Fisher
pMDG, p8.9	Generation of lentivirus	Dr David Darling
pLKO	Generation of lentivirus	AddGene
Retroviral	p-GFP-V-RS	Origene

2.1.9 Cell lines

Cell line	Source	Provider
HEK293-T	Human Embryonic Kidney	Dr David Darling
U-937	Histiocytic Lymphoma	Dr Terry Gaymes
UM-UC3	Urinary bladder	ATCC

2.1.10 cDNA synthesis and qPCR reagent

FastStart Universal Probe Master mix	Roche Diagnostics
MicroAmp™ Fast Optical 96 well reactions plate with Barcode	Life Technologies
Optical cap	Roche Diagnostics
qPCR primers	Integrated DNA technologies
SuperScript® VILO ™ cDNA Synthesis kit	Life Technologies

2.1.11 Western Blot reagents

Spectra Multicolor Broad Range Protein Ladder	Thermo Scientific Fisher
ECL™ Prime Western Blotting Detection Reagent	GE Healthcare
Methanol	Sigma
Dithiothreitol (DTT)	Sigma
Nitrocellulose Membrane	GE Healthcare

Non fat dried milk (Marvel)	Sainsbury
Nupage (4-12%) Bis-Tris gel	Life Technologies
Nupage® LDS Sample buffer (4X)	Life Technologies
Nupage ®Transfer buffer (20X)	Life Technologies
Nupage ®Tris-Acetate SDS Running buffer (20X)	Life Technologies
Tween 20	Sigma

2.1.12 Protein Estimation Reagents

Bovine Serum Albumin (BSA)	Fisher Scientific
Radio Immunoprecipitation Assay (RIPA) lysis buffer	BioRad
Phenyl methane sulfonyl fluoride (PMSF)	BioRad
Sodium orthovanadate	BioRad
Protease Inhibitor cocktail solution	Bio-Rad
Alkaline Copper Tartarate solution	Bio-Rad
Dilute Fooling Reagent	Bio-Rad
Solution S	Bio-Rad

2.1.13 Cell viability and cell cycle reagents

7-Aminoactinomycin D (7-AAD)	BioLegend
Cell based Assay Annexin V Binding buffer	Cayman
Fluorescein isothiocyanate (FITC)	Sigma
FITC Annexin V	BioLegend
Propidium Iodide (PI)	Sigma
RNase A	Sigma

2.1.14 Immunocytochemistry

DAPI (4', 6-diamidino-2-phenylindole)	Sigma
Goat serum	Sigma
Mounting medium	Merck
4% Paraformaldehyde	Sigma

2.1.15 Drugs used

BMN-673	Biomarin
Cisplatin	Sigma
AZD 1152 –HQPA Aurora-kinase B inhibitor	Sigma

2.1.16 Buffers, solutions and media

Blocking solution for western blots: 5% w/v marvel milk powder was dissolved in PBST

Blocking solution for Immunocytochemistry: 1% (w/v) BSA was dissolved in PBS.

Cell cycle staining solution: 400µl of propidium iodide, 50µl of FITC and 100µl of 1 mg/ml RNAse A. Made upto 10ml in PBS.

10X DNA loading dye: 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol, 30% (v/v) glycerol. Made up in dH_2O and stored at room temperature.

Eagle's Minimum Essential Medium: 10% heat inactivated Fetal Bovine Serum (FBS), 100 U/ml penicillin and 100 µg/ml streptomycin was added and stored at 4°C.

2X HEBS: 0.28M NaCl, 0.05M HEPES, 1.5mM Na_2HPO_4, pH 7.0. Made up in dH_2O, filter sterilized and stored at 4°C.

Freezing medium for cell culture: 10% DMSO, 90% FBS. Made prior to use and stored at -20°C.

Luria-Bertani (LB) Broth: 1% (w/v) bactotrytone, 0.5% (w/v) yeast extract, 1% NaCl, 2 ml NaOH for 1 litre LB.

Luria-Bertani (LB) plates containing selective antibiotics: 1.5% (w/v) agar was added to LB broth and then autoclaved. The solution was cooled to 50°C followed by the addition of appropriate antibiotic and then mixed in a hood. The antibiotic containing LB agar was then poured into 100 mm dishes allowed to solidify and then stored at 4°C until further use.

Penicillin/ Streptomycin: The 100X stock solution was aliquoted into 5 ml falcon tubes.

2.2 Protein Extraction

Whole cell protein extracts were prepared using Radio-Immuno Precipitation Assay (RIPA) with cell lysis buffer (50mM Tris-HCl pH 8, 150mM NaCl, 1% NP-40, 1% Sodium deoxycholate, 0.1% SDS) containing 1mM PMSF (phenyl methyl sulphonyl fluoride), sodium orthovanadate and cocktail protease inhibitors (Bio-Rad). 100µl of freshly prepared ice-cold RIPA lysis buffer was used to lyse the cell pellet and was incubated on ice for 30min with intermittent vortexing. Cellular debris was pelleted (7500 × g, 5min, 4°C) and the protein supernatant was transferred to a fresh tube. Protein lysate was stored at -20°C.

2.2.1 Protein Assay

Protein concentration was measured using the DC^{TM} protein assay (Bio-Rad), a detergent compatible colorimetric assay following detergent solubilisation. Based

on the well-documented Lowry method, the assay is based on the reaction of the protein with an alkaline copper tartarate solution and Folins reagent. Copper tartarate treated protein causes the reduction of Folins reagent by loss of oxygen atoms that results in a quantifiable blue colour change proportional to protein concentration. Comparison to a standard curve enables the protein concentration to be measured. Reaction was performed according to the manufacturer's microplate assay protocol. Serial dilution of Protein standard II (bovine serum albumin) 1.4 – 0mg/ml was used to create a standard curve. 5µl of protein lysate or standard was first mixed with 25µl of reagent A (Bio-Rad) and then 200µl of reagent B (Bio-Rad) followed by incubation at room temperature for 15 min. Colour change was measured by absorbance at 750nm and protein concentration calculated against the standard curve.

2.2.2 Western Blotting

Western Blot was used to identify specific proteins in cell lysates under denaturing and reducing conditions. Polyacrylamide gel electrophoresis (PAGE) was used to size separate protein before transfer and immobilisation onto nitrocellulose membrane. The membrane was blocked in 5% milk powder made in PBS Tween for 1 hr followed by probing with a primary antibody specific to the protein under investigation overnight on a roller at 4°C. Detection was achieved using secondary antibody conjugated with horseradish peroxidase (HRP) and enhanced chemoluminescence reagents a mixture of luminol and hydrogen peroxide. In the presence of hydrogen peroxide HRP catalyses oxidation of luminol exciting it to emit light, which can be detected using X-ray, film.

Pre cast gels (4-12% Bis-Tris Gel) (Invitrogen) were used to separate proteins. Protein samples (15 µg) were prepared by adding sample buffer to a concentration

of 1X and 1µl (10%) of DTT (dithiothreitol) followed by heat denaturation at 100°C for 5-10 min. 10µl of sample was loaded alongside a size standard protein ladder and gel were run at 170V for 1hr in 1X running buffer (25mM Tris, 0.2M glycine, 0.1% SDS). Post electrophoresis the stacking gel was trimmed from the lower resolving gel and discarded. To transfer the proteins from the gel a corresponding size of nitrocellulose membrane (GE Healthcare) was placed on top of the gel, which was then sandwiched between, two filter cards of same size one on each side and three nylon pads soaked in running buffer. Extra care was taken not to create or avoid any bubbles between the layers. Proteins were transferred from the gel onto the membrane at 20V in 1X transfer buffer for 1hr. After 1hr the nylon pads were again soaked in transfer buffer and transfer was continued again for 1hr. After the transfer was complete the membrane was incubated in blocking solution (milk powder) for 1hour to reduce nonspecific antibody binding after which an appropriate dilution of primary antibody was made in blocking solution and added to the membrane. The membrane was kept on roller at 4°C overnight. Next day the membrane was washed three times in 0.1% PBST (Phosphate Buffer Saline Tween) five min each before incubating in an appropriate concentration of secondary antibody again made in blocking solution for 1hr at room temperature. After 1hr the membrane was again washed three times in PBST before developing with 5 ml of ECL solution (GE Healthcare) for 5min. The membrane was wrapped in plastic and secured in a developing cassette for X-ray film exposure.

2.3 RNA Extraction

1×10^6 cells were pelleted and resuspended in 1ml TRIzol® (Invitrogen) followed by vigorous vortexing. TRIzol® reagent is a monophasic solution of phenol, guanidine isothiocyanate and other proprietary components which facilitate the isolation of

RNA. TRIzol® reagent maintains the integrity of the RNA due to the highly effective inhibition of RNase activity while disrupting cells and dissolving cell components during sample homogenisation. The pellets were either stored at -80°C or the samples were processed immediately by incubating for 5min at room temperature. 20 µl of chloroform were added to the homogenised samples and then centrifuged at 12,000×g for 15min at 4°C separating the homogenate into a clear upper aqueous layer (containing RNA), an interphase and a red lower organic layer containing the DNA and proteins. The separated aqueous layer was transferred to a fresh microcentrifuge tube, 500µl of 2-propanol was added to precipitate RNA and then the mixture was incubated at room temperature for 10min. The samples were centrifuged at 12,000×g for 10 min at 4°C. The supernatant was discarded and the pellet was washed with 1ml of 75% (v/v) ethanol vortexed and then centrifuged again at 7500×g for 5 min. The supernatant was discarded and the RNA pellet was air-dried for 30 min. The pellet was resuspended in 20 µl of RNase free water and the concentration was determined using a spectrophotometer (Thermo Scientific™ Nanodrop).

SuperScript VILO™ cDNA synthesis kit (Cat No: 117540450) was used to reverse transcribe 100-1000 ng of total RNA according to manufacturer's protocol (Table 2.1). Incubation was carried on the above mix at 25°C for 10 min and then at 42°C for 60min. The reaction was terminated for 5min at 85°C. The cDNA samples were stored at -20°C. The cDNA samples were diluted to 10ng accordingly depending on the starting concentration.

Table: 2.1 cDNA synthesis master mix

Reagent	Volume
10X Superscript Master Mix	2µl
RNA (upto 1µg)	
Water	to 10µl

2.3.1 Quantitative Real Time PCR (q-RT PCR)

FastStart Universal Probe Master mix and probes from Universal Probe Library (Roche Diagnostics) were used to perform Quantitative Real time PCR. The probe finder assay design software available on Roche website was used to design target specific primer sequences and the matching universal library probe. Each sample was run in triplicate to avoid variations (Table 2.2).

Table 2.2: Preparation of Real-time PCR master mix

Reagent	Volume
FastStart Universal Probe Mastermix	10µl
cDNA	5µl
Forward and Reverse Primer (20µM)	1µl
Probe	0.25µl
dH_2O	20µl

The thermo cycling conditions on the StepOnePlus Real-Time PCR (Applied Bio systems) (Table 2.3).

Table 2.3: qPCR Cycling Conditions

Temperature	Time	No of cycles
50°C	2 min	1 cycle
95°C	10 min	1 cycle
95°C 60°C	15 sec 1 min	40 cycles

The $2^{-\Delta\Delta C}{}_T$ was used to calculate relative changes in gene expression determined from real time quantitative PCR experiments using *GAPDH* and *TUBULIN* as housekeeping genes.
Amount of target= $2^{-\Delta\Delta C}{}_T$
$\Delta\Delta C_T = \Delta C_{T\ target} - \Delta C_{T\ calibrator/wildtype}$
$\Delta C_T = C_{T\ target} - C_{T\ GAPDH\ or\ TUBULIN}$
Where C_T is the threshold cycle.

2.4 Cell culture

2.4.1 Culturing of cell lines

Suspension cell line U-937 was cultured in RPMI medium supplemented with 10% (v/v) FBS and 1% (v/v) penicillin/streptomycin. Cells were maintained at a cell density between 1×10^5-2×10^6 cells /ml and media was replaced every 3-4 days.

Adherent cell line UMUC3 was cultured in UMEM media with 10% (v/v) FBS and 1% (v/v) penicillin/streptomycin in a ratio of 1:4 to 1:10. Media renewal was carried 2 to 3 times per week by discarding the culture media and rinsing the cells with PBS to remove all traces of serum that contain trypsin inhibitor. 5-6ml of TrypLE^{TM} Express (1X) Phenol Red was added to the flask and the flask was left in the incubator for 5-15 min for cells to detach. TrypLE^{TM} Express cleaves peptide bonds on the C terminal sides of lysine and arginine and is a direct replacement for trypsin. Its exceptional purity increases specificity and reduces damage to cells that can be caused by other enzymes present in some trypsin extracts. To neutralize the trypsin UMEM was added in larger volume and the cells were then spun down at 300×g for 5 min. Supernatant was removed and cells were resuspended in desired volume after counting the cells.

2.4.2 Cell counting

Cells were mixed with trypan blue in a 1:10 ratio to distinguish between live and dead cells. Trypan blue is a vital stain that is not absorbed by healthy viable cells. When cells are damaged or dead, trypan blue can enter the cell allowing dead cells to be counted. This method is sometimes referred as dye exclusion method. 50µl of cell suspension was resuspended in 400µl of media and then 50µl of trypan blue was added to it and loaded on the haemocytometer (Neubauer cell counting chamber, depth 0.1µl) to count the cells. Cells that were healthy and not stained

were counted from four corner squares and the following calculation was used to determine concentration of cells per ml.

Cells/ml= Average of cells in four corners× dilution factor $\times 10^4$

2.4.3 Cryopreservation of cells

Adherent cell line after trypsinization as described in section 2.4.1 and suspension cell line were spun at 300 × g for 5 min and then resuspended in 1 ml of freezing media and stored in cryovial tubes (UMEM freezing media for adherent cell line and RPMI freezing media for suspension cell line at a density between 1-5$\times 10^6$ cells /ml). The tubes were placed in cryo freezing container (Nalgene, Mr Frosty) overnight and then transferred into liquid nitrogen next day.

2.5 Cell Cycle Analysis

2.5.1 Cell cycle staining

3×10^5 cells were pelleted at 300 ×g for 5min and then washed with PBS. The cells were then fixed using 1 ml of ice cold 70% ethanol (v/v) while continuously vortexing and the cells were stored at -20°C. For cell cycle analysis the cells were centrifuged again at 300 ×g for 5min and the ethanol was removed. The pellet was air dried for 15min to remove traces of ethanol and then 400µl of staining solution Refer to section 2.1.16 was used to resuspend the pellet followed by incubation for 30 min at 37°C and running the sample on a FACS Canto II machine (Becton-Dickinson).

2.5.2 Cell cycle analysis by flow cytometry

DNA analysis is the second most important application of Flow cytometry (Flow cytometry a basic introduction). By measuring the DNA content of individual cells information, can be obtained about their ploidy, which is of particular relevance in

tumours, and for a population the distribution of cells across the cell cycle. Flowjo software (Flow jo LLC) was used to analyse data. Various gating strategies were used to include only single viable cells eliminate debris, dead cells and clumps to produce a histogram plot that gives the percentage of cells in each phase of cell cycle and to ascertain whether knockdown of a particular gene has any effect on any phase of cell cycle.

2.6 Annexin V Staining

Apoptosis or programmed cell death is a normal physiological process for removal of unwanted cells. In apoptotic cells, the membrane phospholipid phosphatidylserine (PS) is translocated from the inner to the outer leaflet of the plasma membrane, thereby exposing PS to the external cellular environment. Annexin V is a 35-36 kDa Ca^{2+} dependent phospholipid-binding protein with high affinity for PS and binds to expose apoptotic cell surface PS. Annexin V can be conjugated to fluorochromes while retaining its high affinity for PS and thus serve as a sensitive probe for flow cytometric analysis of cells undergoing apoptosis. 7-aminoactinomycin D (7AAD) is a viability staining solution and can penetrate the membrane of dying cells and stain them to discriminate between dead and live cells and in an apoptotic assay it is used to distinguish between apoptotic and necrotic cells.

2.5×10^{5} cells were added to a FACS tube. The cells were washed with 1ml of PBS and then spun down at 500×g for 5min.The supernatant was removed and cells were resuspended in either 50µl of buffer, Annexin V, 7AAD or both followed by incubation for 15min at room temperature in the dark and then topped up with 250µl of 1X FACS buffer before reading the sample on the canto machine.

2.7 Immunocytochemistry

3×10^5cells of log phase STAG2 deficient cells were seeded in a 24 well plate in 1ml of media. 1µl of 100 µM BMN-673, was added to deficient as well as control cells and cells were incubated in the drug for 24hr. After 24hr the cells were spun down at 300 ×g for 5 min and then resuspended in 300µl of PBS. 150µl of cell suspensions were cytospin on a poly-lysine coated glass slides at a speed of 300×g for 5min. The cell pellet was allowed to dry and circled using a Dako-pen. The cells were fixed using 4% paraformaldehyde for 15min and then washed three times each with PBS for 5min. 0.1% Triton X 100 (v/v) in PBS was used to permeabilize the cells for 5 min maximum followed by three washes with PBS for 5min and then blocked using 5% sheep or goat serum in 1% BSA in PBS for half an hour. The slides were washed briefly and then incubated in 50µl of primary antibody (RAD51, phospho-γH2AX) made in 1% BSA in PBS for 60min. The slides were washed twice with PBS for 10 min and then incubated in 200µl of secondary antibody anti-IgG –FITC/TRITC for 60min in the dark and then washed again twice with PBS. The slides were then incubated with a drop of DAPI (50-100µg/µl) diluted 1:1000 in PBS for 5min at room temperature and then was washed again for 10min in PBS. The slides were dried at room temperature for 5min and then were mounted using vectashield.

2.8 siRNA

siRNA SMARTpool (Cat no: 10735) from Thermo Scientific was used to target STAG2. Scrambled siRNA of similar construct was used as a control. The results obtained from using SMARTpool siRNA were validated using siRNA from Santa Cruz Biotechnology (catalogue no sc-62970).

2.9 Retroviral siRNA

Four retroviral shRNA against *STAG2* were obtained from Origene

(Cat no: T1336253, T1336254, T1336255 and T1336256).

2.10 Transfection of cells

2.10.1 Transfection by Nucleofection

After harvesting and counting the cells 2×10^6 cells were resuspended in 100µl nucleofection solution. 1µg of shRNA or 100nM of siRNA was added to the cells. The cell suspension was transferred to a nucleofector cuvette and the cuvette was placed in a nucleofactor device after selecting the appropriate programme for cell type (W-001 for U-937). 500µl of media was added to cuvette immediately and the cell suspension was gently resuspended using a Pasteur pipette and transferred to a 6 well tissue culture plate. The transfected cells were topped up with 1ml media and samples were collected after every 24hr for 3 days.

2.11 Lentivirus short-hair pin RNA (LVshRNA) virus particleproduction

Six shRNAs were used to knockdown *STAG2* of which four GFP labelled *STAG2* lentiviral RNAs (Cat no: V2LHS-198853, V2LHS-207886, V3LHS-391825 and V3LHS-391830) were obtained from Fisher Scientific and two shRNAs (Cat no: 3782 and 1221) were obtained from Addgene. HEK 293T (Human Embryonic Kidney) cells were seeded at a density of 1.1×10^8 cells for 1000cm^2 layer flasks [T1000 flasks, Millipore] using DMEM (10% FCS, 1% Pen/Strep) media and transiently co-transfected with three plasmids pMDG (envelope plasmid encoding VSVG), p8.9 (gag, pol) and the shRNA against *STAG2*.

The pool of plasmids containing pMDG (135 µg), p8.9 (315 µg) and shRNA (450 µg) were diluted to a total volume of 11.25 ml with dH_2O. 11.25 ml of 0.5M $CaCl_2$ was added to the mixture. 22.5 ml of 2× HEBS was added to the mixture dropwise while vortexing and incubated for 30 min. After 30min pools of plasmids were

diluted in 150 ml DMEM added to 293T cells and the cells were incubated for 16 hr at 37°C and 5% CO_2. The media was replaced with fresh DMEM after 24 hr and incubated for 24 hr further. After 48hr the supernatant was collected and filtered through a 0.45µm filter and concentrated at 10,000 × g overnight at 6°C. The supernatant was removed and pellet was air-dried for half an hour and resuspended in 500µl of RPMI-1640 medium and stored at -80°C.

2.11.1 Lentiviral transduction

Cells were seeded at a density of 4×10^5 cells in 1 ml of RPMI in a 24 well plate. Viral particles (12 µl) along with 4µg/ml of polybrene were added. After 48hr the cells were topped with 1 ml of fresh media and 2µg /ml of puromycin was added for selection. After 4 days the cell were transferred into a T-25 flask and maintained for 14 days under puromycin selection.

2.12 Cloning

2.12.1 Transformation in DH5α-T1 *E.coli*

Transformations were performed using chemo-competent DH5α- T1 cells (Life Technologies). The cells were thawed on ice for 4-5min and then 1µl of DNA was added, mixed by flicking the tube and then incubated on ice for 30min. The cells were heat shocked at 42°C for 30 seconds and then immediately placed on ice for 2min. 500µl of S.O.S media (Invitrogen) was added to the cells and then incubated at 37°C for 1hr with continuous shaking for cells to recover. After 1hr 100µl of cells were plated on LB agar plates containing ampicillin (100µg/ml). 50 µl of cells were diluted in 50µl of LB and were plated on a second plate, plates were dried in the hood and then placed upside down in a 37°C incubator.

2.12.2 Miniprep: Alkaline Lysis Method

Pure Yield ™ Plasmid Miniprep system (Promega) was used to isolate plasmid DNA from bacterial colonies. Bacterial colonies were picked from LB plates and inoculated in 5ml LB broth containing appropriate antibiotic. The bacterial cultures were grown at 37°C in Innova 4300 incubator shaker. The turbid culture was pelleted at 2340 × g (Rotanta 460R Hettich centrifuge) for 10min and resuspended in 100µl of PBS. Cells were lysed using 100µl of cell lysis buffer and mixed by inverting the tube several times. The reaction was neutralized by adding 350µl of ice-cold neutralization solution (4-8°C) and mixed thoroughly by inverting. The mixture was centrifuged at full speed in a microcentrifuge for 3min and the supernatant was transferred to a Pure Yield ^{TM} Minicolumn. The minicolumn was placed in a collection tube and centrifuged at 13000rpm for 15secs. The flow-through was discarded and 200µl of Endotoxin Removal Wash was added and centrifuged for 13000 rpm for 15secs. The column was washed with 400µl of column wash solution and again centrifuged at 13000 rpm for 30sec. The DNA was eluted by transferring the minicolumn into a 1.5 ml microcentrifuge tube and incubating at room temperature with 30µl of elution buffer followed by centrifugation at 13000 rpm for 15secs. The eluted DNA was stored at -20°C.

2.12.3 Maxiprep: Gen Elute HP Plasmid Maxi Prep

Maxiprep was used to produce DNA in bulk. A single colony was inoculated in 5ml LB broth containing appropriate antibiotic and was incubated for 5-6 hrs at 37°C as previously mentioned. After 5-6 hours the starter culture was diluted into 150 ml LB broth with appropriate antibiotic and grown overnight. The bacteria was pelleted by centrifugation at 5000 × g for 10 min and supernatant was discarded. The pellet was completely resuspended in 12 ml of Resuspension/RNASE A

solution by pipetting up and down and vortexing. 12 ml of lysis solution was added to the suspension and mixed immediately by gently inverting six to eight times. The mixture was then allowed to stand for 5 min until it became clear and viscous and then neutralized by adding 12 ml of chilled Neutralization solution and inverting again which resulted in the formation of white aggregate. 9 ml of binding solution was added to the white aggregate and the solution was immediately poured into the barrel of the filter syringe and allowed to sit for 5 min. In the meantime a GeneElute HP Maxiprep Binding Column was placed in a 50 ml collection tube. 12 ml of Column Preparation solution was added to the column and spun in a swinging bucket rotor at 3000 × g for 2 min. The eluate was discarded. The filter syringe barrel was held over the binding column and pressed gently to expel half of the clear lysate into the column. The lysate was spun in a swinging bucket rotor at 3000 × g for 2 min. The eluate was discarded and the rest of the eluate was added to the column and process was repeated. The column was washed with 12 ml of Wash Solution1 and spin in a spin bucket rotor at 3000 × g for 2 min. The eluate was discarded. The column was washed with 12 ml of Wash solution 2 and spun in a swinging bucket rotor at 3000 × g for 5 min. The binding column was transferred into a clean 50 ml tube. 3 ml2 of elution solution was added to it and was then centrifuged in a swinging bucket at 1000 × g for 5 min.

2.13 Patients Samples

154 MDS patients were selected for mutational screening for *STAG2*. Patients with MDS seen at Kings College Hospital from June 2004 to June 2011 were enrolled in this study. All patients provided written informed consent in accordance with National Research Ethics Protocol (KCLPR060 PR029).

2.14 DNA Techniques

2.14.1 DNA Extraction

DNA was extracted from patients sample using QIAamp DNA purification kit (Qiagen) as per the manufacturers instructions. Bone marrow mononuclear cells were resuspended in 200µl PBS and 20µl protease and then lysed using 200µl of cell lysis buffer (AL buffer) followed by plus- vortexing. Incubation at 56°C for half an hour was followed by addition of 200µl 75-100% ethanol. The mixture was applied to a QIAamp Mini spin column and washed with 500µl of AW1 and AW2 buffer before eluting the DNA in 100µl EB (Elution buffer).

2.14.2 Determination of DNA concentration

DNA concentration was determined using the Thermo Scientific Nanodrop (ND-8000; Nanodrop) spectrophotometer. The spectrophotometer has an absorbance range from 220-750nm. The absorbance spectrum of DNA was measured at 260nm. 2µl of DNA sample was placed on the nanodrop and light absorbed by the sample was measured against the blank (EB buffer). Absorbance ratio at 260/280 is a good representative of purity. Ratio of less than 1.8 indicates protein contamination while more than 1.95 suggests RNA contamination. DNA samples with absorbance ratio around 1.8 were used for sequencing.

2.14.3 DNA Amplification

REPLI-g Midi Kits (Qiagen) was used to amplify DNA as per the manufacturer's instructions from CD34+ of 154 MDS (Myelodysplastic Syndrome) patients. Briefly, DNA was diluted using 1X T.E (Tris- EDTA) buffer (100ng in 5µl). 2.5µl of the diluted sample was taken in a microcentrifuge tube and 2.5µl of D1 buffer prepared according to (Table 2.4) was added, vortexed and then incubated at room temperature for 3min**.**

Table 2.4: Preparation of D1 and N1 buffer

Components	Buffer D1	Buffer N1
Reconstituted Buffer DLB (500µl H_2O + DLB buffer)	9µl	-
Stop solution	-	12µl
Nuclease Free Water	32µl	68µl
Total Volume	41µl	80µl

Neutralization of the sample was achieved by adding 5µl of buffer N1 (refer table 2.4) followed by brief vortexing and centrifugation. A 40µl mastermix containing the REPLI-g DNA polymerase (1µl), REPLI-g Reaction buffer (29µl) and Nuclease- free water (10µl) was added to 10µl of denatured DNA. An isothermal amplification reaction was carried out at 30°C for 16hrs and the polymerase was then inactivated by heating at 65°C for 3min, yielding approximately 40µg of DNA. The amplified product was stored at -20°C.

2.14.4 Checking Amplified DNA

The amplified DNA was diluted 1 in 50 using 1X TE buffer and then was run on a 2% Tris Acetate EDTA (TAE) agarose gel at 120V for half an hour to check amplification.

2.15 Primer Design

The DNA Sequence of *STAG2* was obtained from the Ensemble website and primers were designed using Prime 3 software, creating a PCR product size ranging between 250-350bp. A buffer region of at least 50bp was allocated on either side of the amplicon to enable accurate sequencing of the target. Universal tags (aUSF and aUSR) were incorporated into the primers at the 5' end. All the primers were ordered from IDT (Integrated DNA Technology). Sequence for all the primers is listed in the appendix.

2.15.1 Primer Testing and PCR optimisation

The lyophilized primers were spun down and DNase, RNase free water was added as recommended by IDT to make a stock solution of 100µM. All primers were tested using unamplified control genomic DNA (10ng/µl) using PCR reaction mixture and programme shown in (Table 2.5) and analysed on 1.5% TAE agarose gel alongside ladder (GeneRuler Thermo Scientific SM0331). Samples were run at 120V for 20min in 1X TAE buffer. Amplified DNA samples from 154 MDS patients were used for Next generation sequencing of the *STAG2* gene. The first round PCR reaction and cycling conditions are shown in Table 2.5.

Table 2.5:(a) PCR reaction mixture (b) PCR Programme

(a)

Reagents	Volume
Promega Go Taq	5µl
Primer (5µM)	0.5µl
DNA (10ng/µl)	1µl
H_2O	3.5µl
Total	10µl

(b)

Cycles	Temperature	Time	Steps
1	94°C	5 min	Enzyme Activation
35	94°C	35 secs	Denaturation
	62°C	40 secs	Primer Annealing
	72°C	45 secs	Elongation
1	72°C	15 min	Inserting the missing

A second round of PCR was performed using (0.9µl) of amplified product from the first round of PCR using primers obtained from Sigma Life Science (0.75µl). These primers were targeted to the aUS primers and had sequences at the 5' and 3' end to hybridise with the DNA capture beads in addition to the forward and reverse strands for sequencing. The primers also had a 10 base long MIDs (Multiplex Identifiers) (Catalogue number: 05144507001) to identify individual patients. The programme used for the second round of PCR is mentioned in (Table 2.6).

Table 2.6: 2nd round of PCR programme for 454 Sequencing

Cycles	Temp	Time	Steps
1	94°C	5min	Enzyme Activation
8	94°C	35secs	Denaturation
	60°C	40secs	Primer Annealing
	72°C	45secs	Elongation
1	72°C	15min	Inserting the missing nucleotides

Post- PCR amplified amplicons were quantified using picogreen dye (Invitrogen) on the Rotor-Gene 6000 Multiplexing System (Corbett Research). All the amplicons per patient were equalized in concentration, based on the picogreen measurement by diluting 1µl of DNA in 50µl of 1X Tris-EDTA buffer (Life Science Technologies). The tubes were then run on the Corbett rotor gene alongside six double stranded DNA concentrations of 2 ng/µl, 1 ng/µl, 0.5 ng/µl, 0.25 ng/µl and 0.125 ng/µl. The data was analysed using the Rotor-gene software. After the concentration was equalised, all the amplicons per patient were pooled together and a further picogreen measurement was performed to equalize the concentration of all the patient samples before pooling.

2.15.2 Gel extraction and purification

The libraries were loaded on a 1.8% TAE agarose gel and electrophoresis was carried out at 160V for 40min. The band was excised from the gel and gel purification was carried out using QIA quick Gel Extraction Kit (Qiagen). Briefly, 500µl of buffer QG was added to the gel slice and was incubated at 50°C until the gel was completely dissolved by vortexing every 2min. 10µl of 3M sodium acetate and 100µl of 100% isopropanol were added to increase the yield of DNA. 400µl of

sample was applied to QIAquick column. The column was washed with 500µl of QG buffer and 750µl of PE buffer to remove traces of agarose and salts etc. The DNA was eluted using 30µl of Buffer EB (Elution Buffer) in DNase, RNase free microcentrifugetubes.

2.15.3 Bead Purification

Agencourt AMPure XP magnetic beads (Beckman Coulter, USA) were used to re-purify the DNA again obtained from gel purification. 1.5µl of beads were added for every 1µl of the sample, vortexed and incubated at room temperature for 10min. The sample was then placed on Magnetic Particle Concentrator (MPC) and the supernatant was discarded. The beads were washed twice with 70% ethanol and then dried out at room temperature for 15min. Beads were resuspended in 40µl of warm 30°C water and left at room temperature for 10min to elute the DNA. The DNA sample was then put back on the Magnetic Particle Concentrator and the supernatant was then collected.

2.15.4 Library Preparation

A third picogreen measurement and gel electrophoresis was carried out to confirm the final concentration. The concentration was then converted to 1×10^7 molecules/µl by adding the appropriate amount of molecular grade water.

2.15.5 Sequencing

The GSFLX Titanium LV emPCR Kit (Lib-A) from Roche Applied Sciences was used to perform the emulsion PCR on purified amplicons. Live and mock amplification mix and the capture bead wash buffer were prepared as indicated in the (Table 2.7 and 2.8).

Table 2.7: Preparation of live amplification mix per capture bead

Reagent	Volume
Molecular Grade Water	1200µl
emPCR additive	1500µl
5X amplification mix	780µl
Amplification primer A or B	230µl
emPCR Enzyme mix	200µl
Ppiase	5µl
Total	3915µl

Table 2:.8: (a) Mock Amplification mix (b) 1X capture bead wash buffer

(a)

Reagent	Volume
5X Mock Amplification Mix	2000µl
Molecular Grade Water	8000µl
Total	10000µl

(b)

Reagent	Volume
10X Capture Bead Wash Buffer	1000µl
Molecular Grade Water	9000µl
Total	10,000µl

DNA capture beads A (with forward primer) and B (with reverse primer) were vortexed and centrifuged at 5000×g for 1min. The supernatant was discarded and beads were washed twice with 1ml of 1X capture bead wash buffer. After three washes 50-60µl of supernatant was left behind to prevent beads from drying.

2.15.6 Emulsion PCR

12µl of the purified library at concentration of 1×10^7 molecules/µl was added to capture beads A and B and kept in the fridge. Two cups of oil emulsions were

installed on the TissueLyzer and shaken for 2min at 28Hz. 5ml of 1X Mock Amplification mix was added to each cup and cups were further shaken for 5min at 28Hz to facilitate micelle formation. Next 3.9ml of Live Amplification mix was added to respective capture beads + DNA library solution. Live Amplification mix A + DNA library + capture beads A were added to one emulsion cup and solution with capture beads B was added to other cup and then each cup was put back on Tissue Lyzer for 5min at 12Hz and then solution from each cup was dispensed into 96 well plate (100µl per well) using multistep dispenser pipette. Emulsion PCR (Table 2.9) (Fig. 2.1) was carried overnight as indicated.

Table 2.9: Emulsion PCR conditions

Cycles	Temperature	Time	Steps
1	94°C	4min	Enzyme activation
50	94°C	30secs	Denaturation
	58°C	4min 30secs	Primer Annealing
	68°C	30secs	Elongation
	10°C	∞	On hold

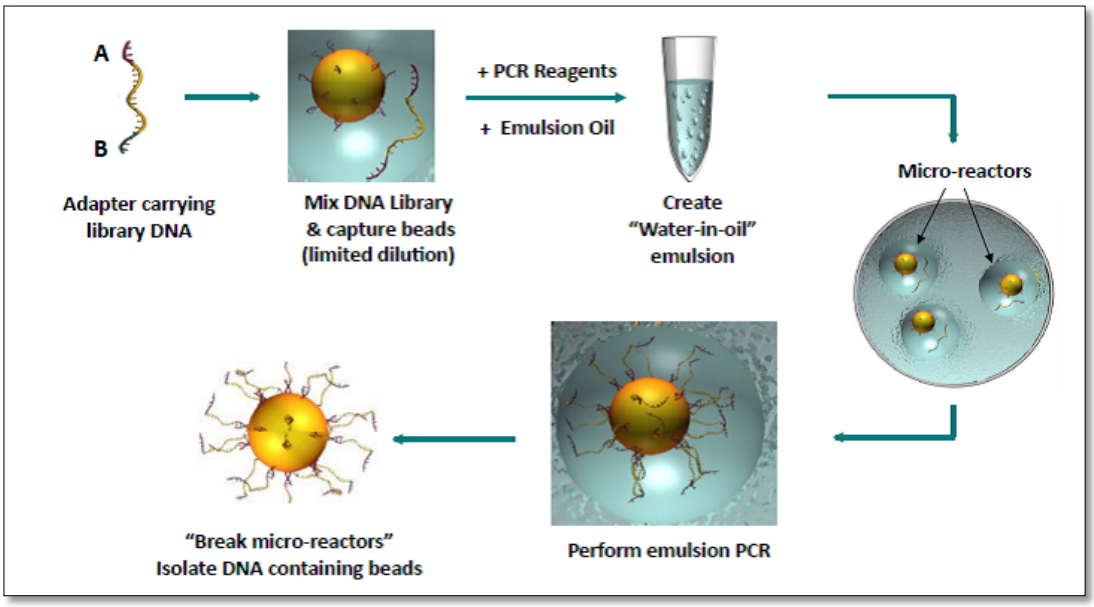


Figure 2.1: Schematic representation of Emulsion PCR. (Image taken from Roche emPCR Method manual Oct 2009).

Adaptors are ligated to DNA fragments for use in subsequent purification, quantification, amplification and sequencing steps by PCR. The library is attached to DNA capture beads. Each bead carries a unique single stranded library fragment. Beads are emulsified with amplification reagents in water and in oil mixture to trap individual beads in amplification microreactors. The entire emulsion is amplified to create millions of clonally copies of each library fragment on each bead. The emulsion is broken down while the amplified fragments remain bound to specific beads.

2.15.7 Vacuum Assisted Emulsion Breaking

Post-PCR emulsions were broken using the GSFLX Titanium emPCR breaking kit and breaking reagents (Lot 93874520). The breaking kit (Fig 2.2) was attached to the vacuum source in a fume hood and the emulsions were rinsed with 3 times 100µl of 100% isopropanol. The capture beads A and B were mixed at this stage and then spun at 1000 × g for 10min. The supernatant was discarded and pellet was washed with 40ml of Enhancing Fluid XT, two times with 100% isopropanol two

times, 100% ethanol and again with Enhancing Fluid by centrifuging at 1000 × g for 5min. The pellet was transferred to microcentrifuge tubes and washed twice with 1ml of Enhancing fluid at 4000 × g for 1min.

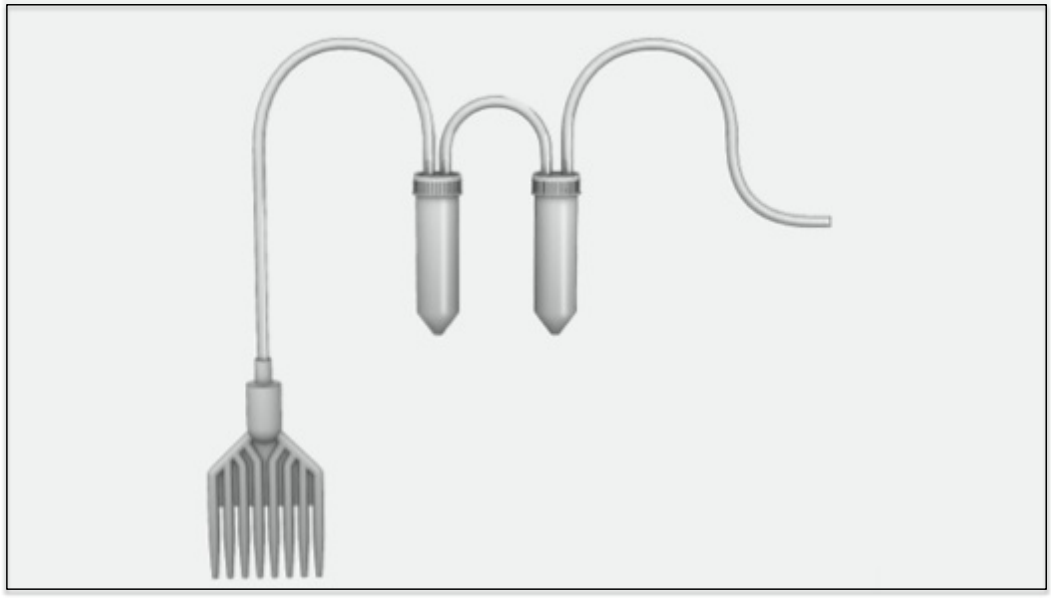


Figure 2.2: Assembled set-up for vacuum assisted emulsion breaking and bead recovery. (Image taken from Roche emPCR Method manual Oct 2009).

2.15.8 DNA Library Bead Enrichment

After the final wash supernatant was removed and 1ml of fresh Melt solution (500µl of 5M NaOH and 19.5ml of Molecular Grade Water) was added, vortexed and incubated for 2min at room temperature. The supernatant was discarded and 1ml of Melt solution was added again. The Melt solution was washed using 1ml of Annealing buffer XT twice and spun at 3200 × g for 2min. The pellet was resuspended in 45µl of annealing buffer and 12.5µl each of enrichment primer A and B was added (from GSFLX Titanium MV emPCR Kit) and kept on a heat block at 65°C for 7-8 min and then was subsequently kept on ice for 2 min to stop the reaction. The pellet was then washed twice with 1ml of enhancing fluid and spun at

3200 × g for 2min to remove residual primers. The pellet was then resuspended in 800µl of Enhancing Fluid (XT).

2.15.9 Preparation of Enrichment Beads

The tubes containing the Enrichment beads were vortexed and placed on a Magnetic Particle Concentrator (MPC). The supernatant was then discarded and beads were washed twice with 1ml of enhancing fluid. After the final wash, 320µl of enhancing fluid was added to the beads and 80µl of enrichment beads were added to the pellet resuspended in enhancing fluid from previous step. The tubes were secured with autoclave tapes and placed on a rotor for 15min, pooled in 15ml glass tubes and placed again on the MPC. The beads were washed with 3-4ml of enhancing fluid 15 times till a clear supernatant was obtained.

2.15.10 Collection of Enrichment Beads

Enriched beads were resuspended in 1.5ml Melt solution (prepared earlier), vortexed and placed back on the MPC and left for 5min. The supernatant was collected and the denaturation step was performed again to get maximum yield. The Melt solution was washed twice with 1ml annealing buffer and centrifuged at 3200 × g for 2min. The pellet was resuspended in 1ml of annealing buffer.

2.15.11 Determine Bead Enrichment

Table 2.10: Coulter Counter Z1 models settings for bead counting

Parameter	Dual Threshold (µm)
Upper Size	25
Lower Size	15
Count Mode	Between
Aperture 100µm Upper Size	100

3µl aliquot of the beads were counted using Beckham Coulter Z1 particle counter.

% Bead Enrichment = (Number of Enriched Beads/35 × 10^6 beads/cup) × 100

2.15.12 Sequencing Primer Annealing

25µl of sequencing primer A [from GSFLX Titanium LV emPCR Kit] + 25µl of sequencing primer B was added to the pellet in annealing buffer and was placed on a heat block at 65°C for 7-8min. The reaction was stopped by placing the tube on ice for 2min. The library was washed three times with 800µl of annealing buffer, spun at 3200 × g for 2min and then was stored in the fridge in 1ml of annealing buffer.

2.15.13 Preparation of Bead Buffer 2 (BB2)

Bead wash buffer (200ml) was placed on ice for 10min. 1.2ml of the Supplement CB and 34µl of Apyrase was added to make the Bead Buffer 2 (BB2) Table 2.11.

Table 2.11: Components of DBIM buffer

BB2(µl)	Polymerase Cofactor (µl)	DNA Polymerase (µl)	Total (µl)
1570	150	300	2020

2.15.14 Preparation of DNA Bead Incubation Mix (DBIM)

Bead Preparation [Titanium Sequencing Kit XLR70]

1. Packaging beads were washed three times with 1ml of BB2 and spun at 10,000 × g for 5min. 550µl of BB2 was added and the beads were left on ice.
2. Enzyme and PPiase beads were vortexed for 30sec each and then placed on the MPC. Each tube was washed three times with 1ml of BB2. Enzyme beads were left in 1ml of BB2 while PPiase beads were left in 500µl of BB2 and left on ice.
3. DNA + Control Beads: The libraries were thawed overnight and the next day, spun and 2 ml of annealing buffer was added. The library was

measured again using the bead counter. The appropriate volume of DNA library beads (2 million beads per region) were separated and 20 µl of DNA control beads was added / region along with 1ml of DBIM/ region and placed on the rotor for 15 min.

2.15.15 Layer Preparation

Layers 1, 3 and 4 were prepared according to table in falcon tubes and placed on ice (Table 2.12).

Table 2.12: Dilution of beads for layer preparation

Bead Layer	BB2 (µl)	Enzyme Beads (µl)	PPiase Beads (µl)	Total (µl)
Layer 1 Enzyme Beads	3250	550	-----	3800
Layer 3 Enzyme Layer	2500	1300	-----	3800
Layer 4 PPiase Beads	3340	-----	460	3800

2.15.16 Assembling the Bead Deposition Device (BDD)

The Roche Pico Titre Plate Kit was washed and assembled as per manufacturers instructions as indicated in the (Fig 2.3) by placing the PTP device onto the BDD base and aligning the notched corner.

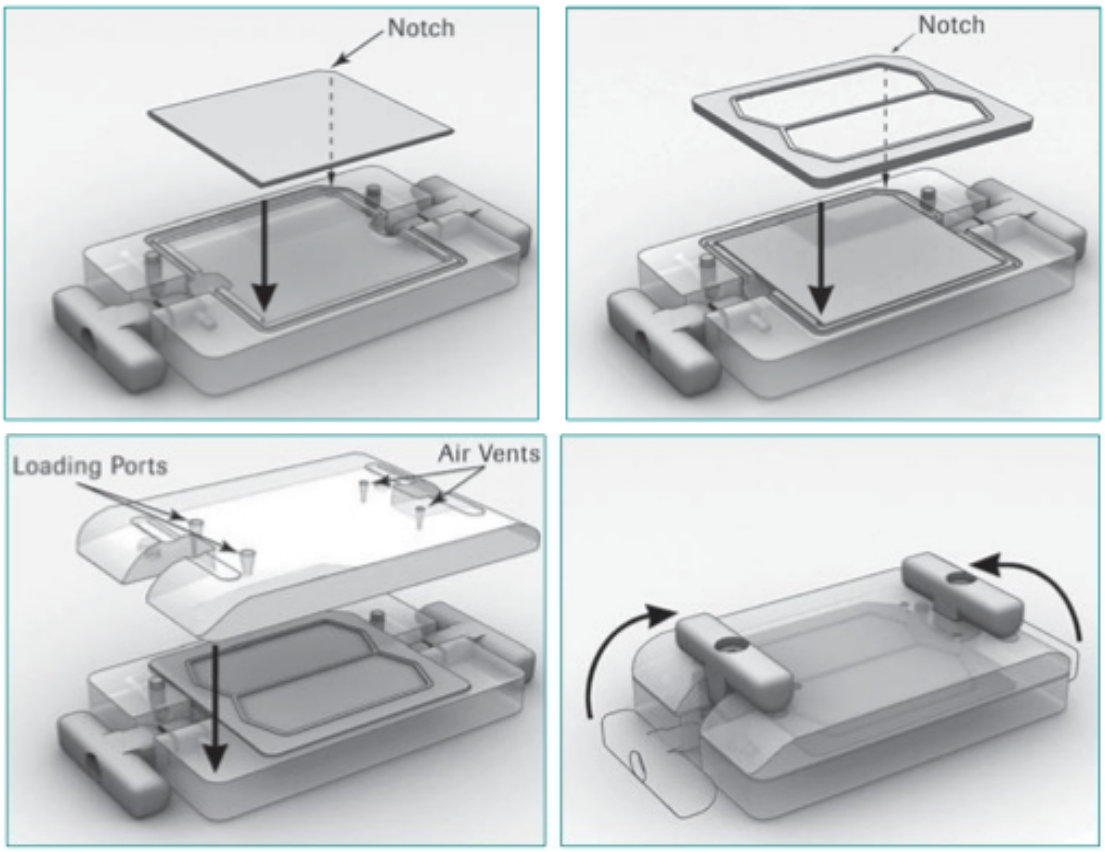


Figure 2.3: Assembling the BDD. (Image taken from Roche sequencing method manual Oct 2009).

1.9ml of each layer was loaded on the PTP plate through the loading ports and centrifuged at 1620 × g for 5min. Previous layer were aspirated before adding subsequent layers. After the last layer was aspirated out the plate was removed from the gasket and loaded on the 454 sequencer.

2.15.17 Run Requirements

One maintenance wash and two pre-washes were carried out before the run. The PTP cartridge along with the camera faceplate were cleaned and the sequencing kit reagent LR70 required for the run were loaded on the machine as per the manufacturer's instructions.

2.16 Sanger Sequencing

Sanger Sequencing was carried out on skin and T cells to confirm that *STAG2* mutations found in six patients were not constitutional polymorphisms. Amplicons in which mutations were detected were amplified using respective primers and the PCR programme described in (Table 2.5). The PCR products were cleaned using ExoSAP-IT from Affymetrix (6µl sample + 3µl ExoSAP). The programme used is described in (Table 2.13).

Table 2.13: Exo-Sap –IT clean-up programme

Time	Temperature	Steps
15min	37°C	ExonucleaseI removes leftover primers
15min	80°C	Shrimp Alkaline Phosphatase removes dNTPs

The concentration of the amplified DNA was measured and 10-20ng of DNA was used for Step PCR to incorporate the dye into the DNA sequence (Table 2.14 and 2.15).

Table 2.14: Step-PCR reaction mixture

Reagents	**Volume (µl)**
5X Sequencing Buffer	2
Big dye	0.5
5µM Universal Primer (aUSF) or (aUSR) each	1
1st Round of PCR product	3
H_20	3.5
Total	11

Table 2.15: PCR Programe

Cycle No	Temperature (°C)	Time	Cycles	Comment
1	95	1min	× 1	Enzyme Activation
2	95	10secs	×15	Denaturation
	55	5secs		Primer Annealing
	60	1.15min		Elongation
3	95	10secs	× 5	Denaturation
	55	5secs		Primer Annealing
	60	1.30secs		Elongation
4	95	10secs	× 5	Denaturation
	55	5secs		Primer Annealing
	60	2min		Elongation
5	4	∞		On hold

Purification of the PCR products was carried out using 45µl of the Big Dye XTerminator and 5µl of the SAM solution (Applied Bio systems). The mixture was vortexed on a plate mixer for 30min followed by centrifugation for 2min at 1000 × g. Sanger Sequencing was carried out using an ABI 3130x genetic analyser and the DNA sequences were visualized for mutations using SeqScape Software (v2.5) and analysed using Sequencing Analysis Software (Applied Bio systems, Foster City CA).

2.17 MiSeq Amplicon Sequencing

The same cohorts of patients used for sequencing of STAG2 were used to sequence other three members of the cohesin complex *SMC1A*, *SMC3*, *RAD21*. The same methodology was applied as for the 454 sequencing up until the first round of PCR. After the first round of PCR, all the amplicons for three genes from four patients were run on 1.5% agarose gel and from the intensity of the bands the concentration of DNA was adjusted. The amplicons for all the genes for all the patients were

pooled together and gel and bead purification was performed as mentioned in section (2.15.2 and 2.15.3) followed by picogreen measurement to obtain a concentration of 0.2 ng/µl for Nextera PCR.

2.17.1 Tagmentation of Input DNA

During this step input DNA was tagmented (tagged and fragmented) by the Nextera XT transposome. The Nextera XT transposome simultaneously fragments the input DNA and adds adaptor sequence to the ends allowing amplification by PCR in subsequent steps.

2.17.2 Preparation

The Amplicon Tagment Mix (ATM), Tagment DNA Buffer and input DNA were removed from -20°C storage and thawed on ice. Neutralize Tagment Buffer (NT) needs to be at room temperature and was visually inspected to ensure that there is no precipitate. If there was precipitate it was vortexed until all particulates were resusupended. After thawing, all the reagents were adequately mixed by gently inverting the tubes 3-5 times followed by a brief spin in a microcentrifuge.

2.17.3 Make NTA

A new 96 well plate was labelled NTA (Nextera XT Tagment Amplicon Plate). 2.5µl of TD buffer was added to each well used in the assay. Tips were changed between samples. 1.25µl of input DNA at 0.2 ng/µl was added to each sample well of the NTA plate. 1.25µl of ATM was added to the wells containing input DNA and TD buffer. The samples were gently pipetted up and down five times to mix using a multichannel pipette. The NTA plate was covered with a microseal and centrifuged at 280 ×g at 20°C for 1min. The NTA plate was placed in a thermocycler and incubated at 55°C for 5min and then held at 10°C. Once the sample reached 10°C NTA was immediately neutralized.

2.17.4 NTA Neutralization

The microseal was removed carefully and 1.25µl of NT Buffer was added to each well of the NTA plate. Using a multichannel pipette, samples were pipetted up and down five times to mix. The NTA plate was covered with microseal, centrifuged at 280 ×g at 20°C for 1min and then placed at room temperature for 5 min.

2.17.5 PCR Amplification

In this step the tagmented DNA was amplified via a limited cycle PCR program. The PCR step also added index 1 (i7) and index 2 (i5) and sequences required for cluster formation (Lot no 9855502). The Nextera PCR Master Mix (NPM) and the index primers were removed from -20°C storage and thawed on bench at room temperature for approximately 20min. Once thawed the tubes were gently inverted 3-5 times to mix and briefly centrifuged in a microcentrifuge. The index primers were arranged in the TruSeq Index Plate Fixture using the following arrangement: Index 1 (i7) primer tubes (orange caps) in order horizontally, so that N701 is in column 1 and N712 is in column12. Index 2 (i5) primers (white caps) in order vertically, so that S501 is in row A and S508 is in row H (Fig 2.4).

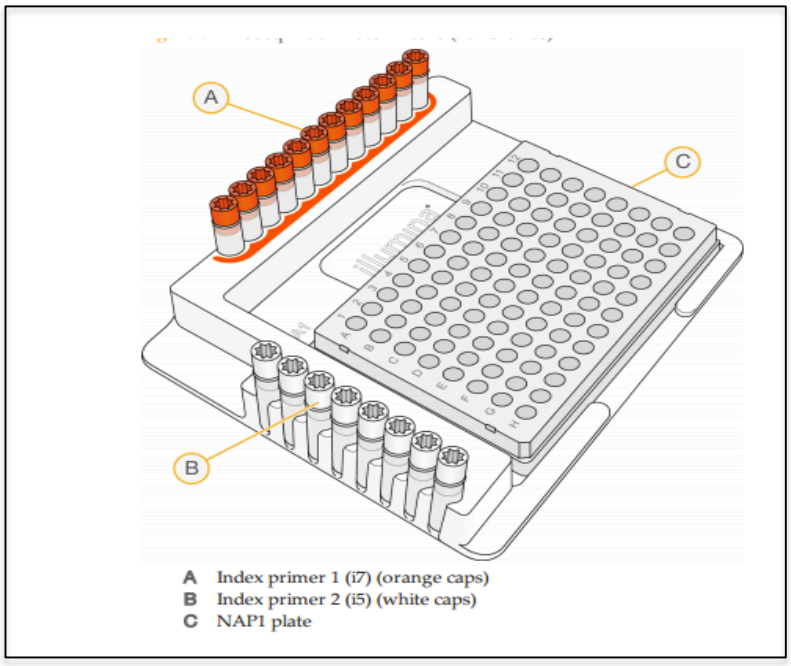


Figure 2.4: True Index plate fixture arrangement. (A) Rows A-D Index 2(i5) adapters. (B) Column 1-6 Index 1 (i7) adapters. (C) Hard shell PCR plate. (Image taken from Nextera XT sample preparation guide)

2.17.6 Amplify NTA

The NTA plate was placed in the TruSeq Index Plate Fixture. 3.75µl of NPM was added to each well of the NTA plate containing index primers. Tips were changed between samples. Using a multichannel pipette 1.25µl of Index 2 primers (white caps) was added to each column of the NTA plate. Tips were changed between the columns to avoid cross contamination. Using a multichannel pipette 1.25µl of Index 1 primers (orange caps) were added to each row of the NTA plate. Tips were changed again after each row to avoid index cross contamination. Using a multichannel pipette, samples were pipetted up and down 3 to 5 times to mix and again tips were changed between samples to avoid index and sample cross contamination. The plate was sealed with a microseal and centrifuged at 280 ×g at 20°C for 1min. PCR was performed using the following program on a thermal cycler (Table 2.16).

Table 2.16: Thermal cycler conditions used for Nextera -PCR

Cycles	Temperature	Time
	72°C	3min
	95°C	30sec
12	95°C	10sec
	55°C	30sec
	72°C	30sec
	72°C	5min
Hold at 10°C		

2.17.7 PCR Clean Up

AMPure XP beads were used to purify the library DNA and provide a size selection step to remove very short library fragments from the population. All patient samples post Nextera amplification was quantified using the picogreen dye as described previously in section 2.16.1 patient libraries were pooled together in equal concentration. Pooled patient libraries were purified using the Agencourt AMpure XP beads as described in section 2.16.3 and quantified again using the picogreen dye.

2.18 MiSeq Amplicon Sequencing

2.18.1 Preparation of HT1 buffer

HT1 (Hybridization Buffer) (Lot no 9853711) was used to dilute denatured libraries before loading libraries onto the reagent cartridge for sequencing. The tube of HT1 was removed from -20°C storage and set aside at room temperature to thaw. Freshly diluted NaOH was used to dilute libraries for cluster generation on MiSeq. A fresh dilution of 0.2N NaOH was prepared by combining 800µl of Laboratory-grade water with 200µl of 1.0 N NaOH of in a microcentrifuge tube.

2.18.2 Denature DNA for 4nM Library

5µl of 4nM sample DNA and 5µl of freshly prepared 0.2N NaOH were combined in a microcentrifuge tube. The rest of the NaOH was used to make a PhiX control (Lot no 9702818). The sample solution was vortexed briefly to mix and then centrifuged at 280 × g for 1min. The sample was incubated for 5mins at room temperature to denature the DNA into single strands. 990µl of pre-chilled HT1 was added to the tube containing 10µl denatured DNA.

The result is a 20pM denatured library in 1mM NaOH. The denatured DNA was placed on ice prior to proceeding to the final dilution. The 20pM DNA was diluted further to give a final concentration of 11pM in 600µl. The tube was inverted several times to mix and then pulse centrifuged. The denatured and diluted DNA was placed on ice until ready to load the samples onto MiSeq reagent cartridge (Lot No 9853711).

2.18.3 Preparing PhiX Control

2µl of 10 nM PhiX library was combined with 3µl of 10mM Tris-Cl pH 8.5 with 0.1% Tween 20 to dilute PhiX library to 4nM. 5µl of 4nM PhiX was combined with 5µl 0.2 N NaOH and vortexed briefly to mix the 2nM PhiX library solution. The template solution was centrifuged to 280 × g for 1min and was incubated for 5min at room temperature to denature the PhiX library into single strands. 10 µl of denatured PhiX library and 990µl of pre-chilled HT1 was added to result in a 20pM PhiX library. 6µl of denatured and diluted PhiX control was combined with 594µl of denatured and diluted sample library.

2.18.4 Load Sample Libraries

The foil seal covering the reservoir labelled load samples of the reagent cartridge was cleaned using a low lint lab tissue. 600µl of prepared library was pipetted into

the load sample reservoir. Using forceps, the flow cell was immersed in a storage buffer in a flow cell container and lightly rinsed with laboratory grade water to remove excess salts and then dried with a lint free lens cleaning tissue. The flow cell glass was cleaned with alcohol wipe to make sure that the glass was free of streaks, fingerprints and lint or tissue fibres and was loaded into the flow cell stage. The PR2 bottle was removed from 4°C, inverted, the lid removed and was then placed in the reagent compartment door. The reagent cartridge was placed in the reagent chiller door. After loading, the flow cell, the reagents and the run parameters were reviewed and a pre-run check was performed before starting the run. When all the items successfully passed the pre-run check, the actual run was performed and the run was monitored using sequencing analysis viewer (SAV) in BaseSpace. An instrument wash was performed after completion of the sequencing run.

2.18.5 Data Analysis

The MiSeq data analysis pipeline was used to analyse the data. *.bcl files were generated during the sequencing run by the Illumina RNASeq 2000 Real Time Analysis (RTA) software. fastqfiles were generated from *.bcl files. Sequence alignment to the human target reference genes (NCBI37/Hg19) and variant calling for SNPs/Indels was performed via Burrows-Wheelers aligner (BWA) GATK pipeline (Li and Durbin, 2009, McKenna et al., 2010). Illumina Variant studio was used to visualize processed VCF and BAM files. Validation was performed on variants if not found in dbSNP142 and 1000 genome database at > 0.01 population frequency.

Chapter Three

MUTATIONAL ANALYSIS OF COHESIN GENES IN MDS

3.1 Introduction

The Myelodysplastic Syndromes are a group of clonal heterogenous disorders characterized by inefficient haemopoiesis, hypercellular bone marrow, dysplasia of blood cells and peripheral blood cytopenias. High-throughput second generation sequencing, (SGS) has elucidated a comprehensive biomarker mutation profile leading to an increased diagnostic sensitivity, specificity and improved prognostic risk stratification except in the context of CHIP (Clonal hematopoiesis of Indeterminate potential) and CCUS (Clonal cytopenia of undetermined significance) where sensitivity of this technique is not clear. These entities are not specific disorders per se rather, they are general terms that can be used to describe patients without a specific diagnosis in whom cytopenias or clonal mutations represent a possible sign of disease that may manifest more clearly in the future.

Infinitely deep sequencing will detect acquired mutations in every adult so proposals have been made to distinguish between the above terms on the basis of variant allele frequency (VAF). VAF is the proportion of sequencing reads with a mutation which in the absence of a copy number variant is roughly parallel to the size of the clone bearing the mutation. With good natural history studies, we may learn that specific mutations and variant allele frequencies have a similar natural history to MDS diagnosed by conventional means, thus allowing diagnosis of MDS without the morphologic dysplasia that is currently required for the diagnosis in the absence of karyotypic abnormalities.

DNA sequencing has come a long way since the days of two dimensional chromatography in the 1970s (www.Illumina.com). Genome sequencing has been revolutionised since the completion of the human genome project (HGP) and has

moved away from first generation sequencing approaches to whole-genome sequencing (WGS) with changes in the accompanying software for denovo assembly of large scale sequencing data. Highly streamlined sample preparation steps prior to DNA sequencing in second generation sequencing (SGS) offers significantly increased time saving and a minimal requirement for associated equipment in comparison to the highly automated multistep pipelines necessary for first generation sequencing (Table 3.1). Instead of sequencing a single DNA fragment SGS extends this process across millions of fragments in a massively parallel fashion. The sequencing library is prepared by random fragmentation of DNA followed by 5' and 3'adapter ligation followed by PCR amplification. These universal adapters are specific to each sequencing platform that can be used to polymerase amplify the fragments during specific steps of the protocol. Each fragment is then amplified into distinct clonal clusters. During clonal cluster formation the sequencing template is immobilized on a proprietary flow cell surface (Illumina sequencing) or a bead (454 Next generation sequencing) designed to present DNA in a manner that facilitates easy access to enzyme while ensuring high stability of surface bound template and low non specific binding of fluorescently labelled nucleotides and resulting in thousands of identical copies of each single template molecule in close approximation. During each cycle the nucleotides are identified by fluorophore excitation through the incorporation of fluorescently labelled deoxynucleotide triphosphates (dNTPs) into a DNA template strand during sequential cycles of DNA synthesis catalysed by DNA polymerase.

Table 3.1: Advantages of SGS. Advantages and mechanism of first and second generation sequencers (Liu et al., 2012).

Sequencer	MiSeq	454 GS FLX+	Sanger 3730xl
Sequencing mechanism	Sequencing by synthesis (SBS)	Pyrosequencing	Dideoxy chain termination
Read length (bp) (No of bases sequenced)	Depends on the MiSeq kit used 2×250bp using MiSeq Reagent Kit v2 2×300bp using MiSeq Reagent Kit v3	700bp	400-900 bp
Accuracy	Error rate below 0.4%	99.9%	99.999%
Reads	15 million reads	1million	-
Output data/run	120 Mb-1.5 Gb	0.7 Gb	1.9 ~84 Kb
Time/run	4 hrs to 55 hrs depending on the number of cycles performed	24 hrs	20 min ~ 3 hrs
Advantage	Moderate cost instrument and runs. Low cost per Mb for a small platform. Fastest Illumina run times and longest Illumina read length	Read length, fast	High quality, long read length
Disadvantage	Major concentrate in flowcell surface size, insert sizes, and how to pack cluster in tighter	Error rate with polybase more than 6, high cost, low throughput	High cost, low throughput

3.2 Aim

Large scale genomic sequencing has identified mutations in cohesin complex in a large variety of myeloid malignancies particularly in AML (10-20%) (Lindsley et al., 2015, Thol et al., 2014, Thota et al., 2014). Mutations have also been reported in various types of solid cancers Ewing's sarcoma (20%) bladder cancer (20%), (Balbas-Martinez et al., 2013, Guo et al., 2013, Solomon et al., 2013, Solomon et al., 2011) Apart from somatic mutations germline mutations in cohesin complex has also been associated with developmental disorders like Cornelia de Lange Syndrome (CdLS) and Roberts syndrome characterized by developmental disorders, intellectual disability. (Mannini et al., 2013).

Mutations in cohesin complex fall into two categories (1) nonsense and frameshift mutations are found in *STAG2* and *RAD21* genes (2) missense mutations are mostly observed in *SMC1A* and *SMC3* genes (Kon et al., 2013). Mutations in different members of the cohesin complex appear to be mutually exclusive which disapproves functional redundancy between these proteins however how these mutations contribute towards pathogenesis of disease is still undermined. To unravel the role of cohesin mutations in myelodysplastic syndrome second-generation sequencing was carried out in a cohort of 154 MDS patients. The main aim was to

(1) Determine the prevalence and allele burden of mutations affecting the cohesin complex in MDS patients.

(2) Correlate cohesin mutations with other mutations, response to therapy, survival and karyotype of MDS patients.

Table 3.2: Demographics, Clinical characteristics and therapy in 154 MDS patients. Cytogenetic failed in 3 patients † denotes other treatments which include; lenalidomide, thalidomide, cyclosporine and Anti-thymocyte globulin * includes 13 patients who had HSCT (Haematopoietic stem cell transplant) after receiving 5-azacitidine. [n- Represents number of patients] % Represents percentage of the patients.

Patient Characteristics	Overall
Total Patients	154
Age, years	
Median	65.5
Range	17-85
Sex	
Male [n (%)]	104 (67%)
Female [n (%)]	50 (32.4%)
WHO category*	
RA/RCMD [n (%)]	40 (26%)
RARS/RCMD-RS [n (%)]	24 (16%)
RAEB1/2 [n (%)]	49 (32%)
sAML [n (%)]	15 (10%)
tMDS/AML [n (%)]	12 (8%)
CMML & MPD/MDS-U [n (%)]	14 (9%)
Bone Marrow Blasts	
Median (%)	5
Range	0-80
IPSS cytogenetic risk group	151
Good	90
Intermediate	17
Poor	44
Median follow up (months)	21.4
Transfusion dependency	
Yes	80 (51%)
Progression to AML	
Yes	44 (28%)
Treatments	
5-Azacytidine	66
Other Treatments †	14
HSCT*	35
No Treatments	39

3.2.1 Testing for *STAG2* gene

REPLI-g Midi Kits from Qiagen (refer to section 2.14.1 for more details) was used to amplify patient DNA isolated from bone marrow total nucleated cells to increase the yield of DNA as cell number were low. Unamplified patient gDNA was used for mutation confirmation in an independent experiment. The amplified DNA was diluted using TE buffer and 2µl was run on a 1.8% agarose gel to check for amplification. However, bands were detected in the negative control (water instead of DNA was used as negative control) (Fig 3.1a). Although the Go Taq Polymerase used to perform the PCR reaction is a high fidelity polymerase but despite that there is still a low rate of misincorporation of nucleotides that might have been the reason for the amplification product observed in the negative control. To check whether the amplified products were real or artifactual, PCR was performed for two positive and three negative controls for amplicon 1 of *STAG2* gene and then were run on a 1.8% TAE agarose gel. No bands for the negative samples were found. PCR was again performed with two negatives and two positive samples for all the amplicons of *STAG2* gene and again run on 1.8% TAE gel. Bands for amplicon 9 in one negative sample and amplicon 10 in another negative sample for *STAG2* gene were detected. Primers for amplicons 9 and 10 were thus redesigned using the http://bioinfo.ut.ee/primer3/ software and PCR reactions were re-performed with the newly designed primers. No bands were detected in the negative control (Fig 3.1b).

In order to determine the primer efficiency PCR reactions were initially performed using 5µM of primer and 2µl of 10ng/µl of unamplified control genomic DNA. Two PCR runs were performed using all the STAG2 primers and then the PCR products were loaded on 1.5% TAE agarose gels. Due to PCR non amplification of the

amplicons primers 8, 10, 14 did not amplify a PCR product (Fig 3.1c) thus a gradient PCR for each primer at three different annealing temperatures 56°C, 58°C and 60°C was performed (Fig 3.1d). Upon altering the annealing temperatures PCR products were obtained at the correct sizes at all temperatures and thus an annealing temperature of 60°C was selected in all reactions. The PCR product bands were of the same intensity and corresponded to the correct product size.

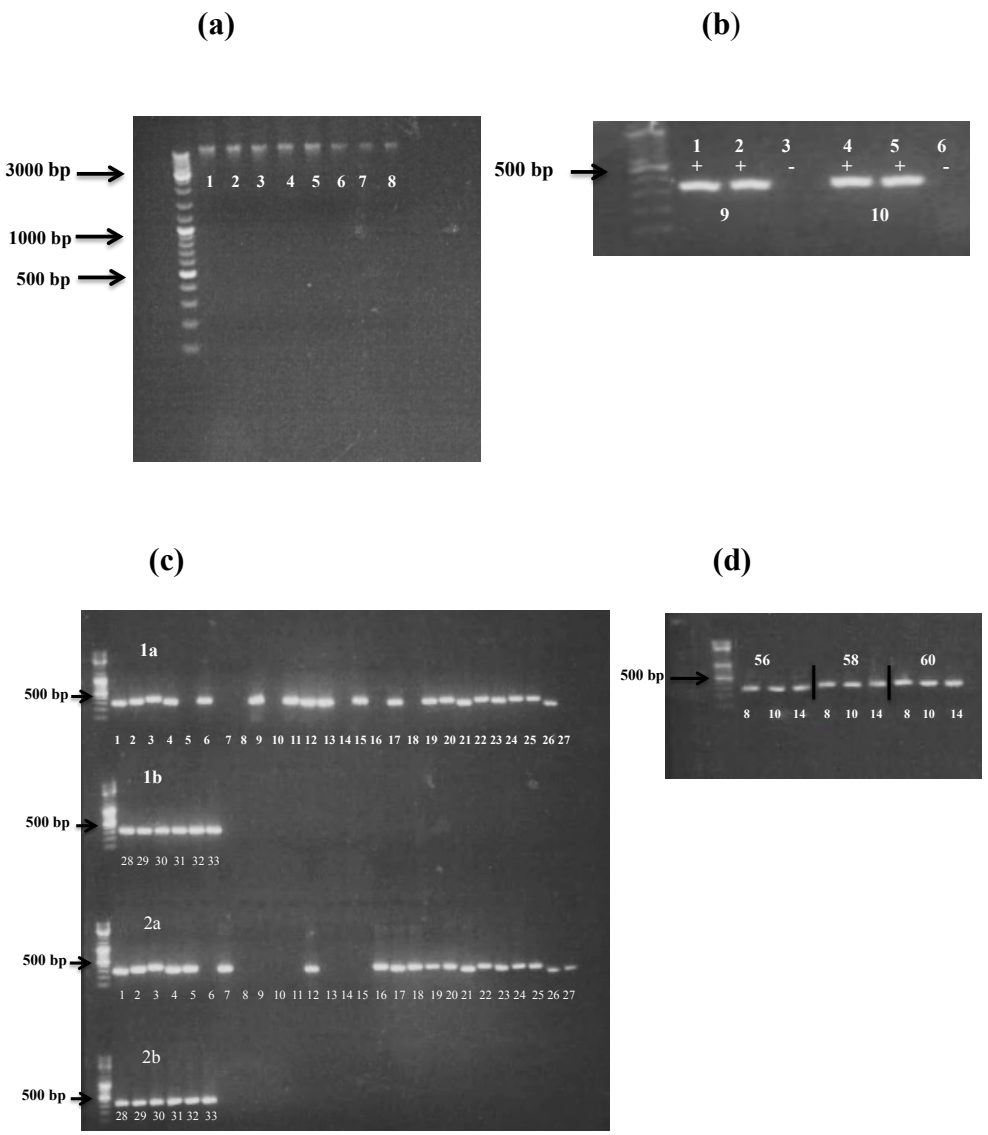


Figure 3.1: DNA amplification of *STAG2* using Qiagen REPLI-g kit and Primer Optimization. (a) Lane 1-5 amplified DNA, Lane 6-8 negative controls. (b) Newly designed primers for amplicon 9 and 10. Row 1-3 two positive and one negative samples tested with newly designed primer for amplicon 9. Lane 4-6 two positive and one negative sample tested with newly designed primer for amplicon 10. No bands were detected in the negative control. (c) The primers were tested twice on unamplified genomic DNA. Row 1a-1b All the 33 primers were tested on unamplified genomic DNA. Row 2a-2b PCR was performed again with unamplified DNA for all the amplicons. On comparison it was found that primers for amplicons 8, 10 and 14 had not worked and hence gradient PCR at three different annealing temperatures was performed. (d) Gradient PCR was performed using primers 8, 10, 14 and tested at three different annealing temperatures 56°C, 58°C and 60°C. Primers worked at all the three different annealing temperatures and hence 60°C was selected as the annealing temperature.

3.2.2 Next Generation Sequencing

Amplified DNA from all the 154 samples was used to identify mutations in the *STAG2* gene using the Roche 454 GS FLX $^{+}$ second generation sequencing platform which offers read length upto 1kb. 1st round of PCR was performed using the optimised primers (Refer to section 2.15.1 for more details) and then two random patient samples from each amplicon were analysed on 1.8% TAE agarose gel. The bands were all of the same intensity and corresponded to the anticipated product size Fig (3.2). 0.9 µl of DNA sample from all amplicons were used for the 2nd round of PCR (Refer to section 2.15.1 for more details). Again two random samples from each amplicon were visualised on agarose gel after 2nd round of PCR (Fig 3.3).

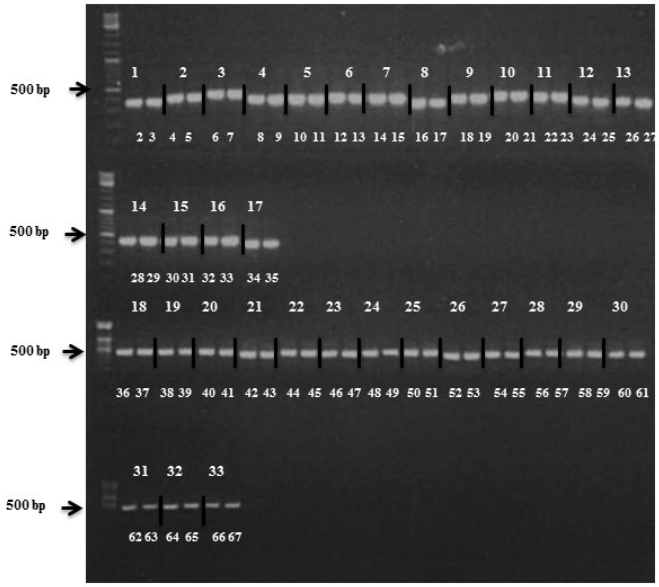


Figure 3.2: Visualization of two patient samples per amplicon after 1st round of PCR. Lane 1-33 upper section represent number of exons while lane 2-67 lower section represent two patient samples per amplicon.

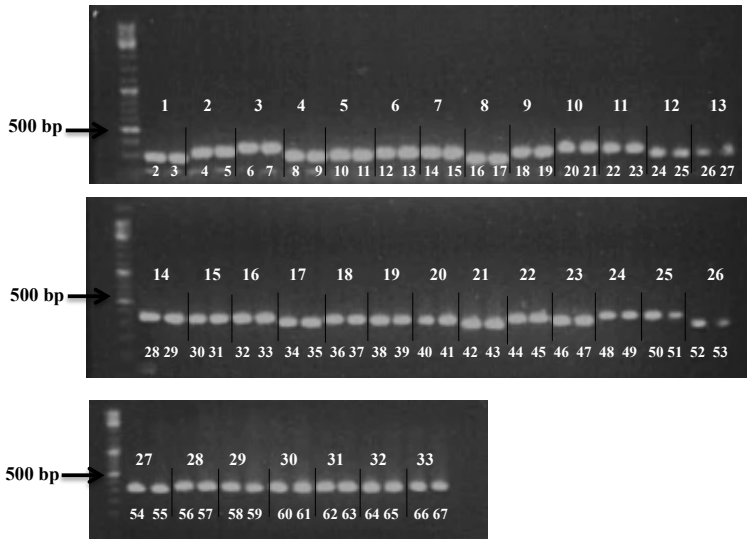


Figure 3.3: Visualization of two patient samples per amplicon after the 2nd round of PCR for *STAG2*. Lane 1-33 upper section represent number of exons while lane 2-67 lower section represent two patient samples per amplicon.

After the second round of PCR all the 33 amplicons were quantified using Quant-iT picogreen dsDNA Assay kit on the Rotor–Gene 6000 Multiplexing System. This measurement was performed to equalize the concentrations of all the amplicons per patient and then all the amplicons for each patient were pooled together. Following on picogreen measurement was again performed to quantify patients samples and then all patients were pooled together in equimolar concentrations (Refer to section 2.18.1 for more details). The final mix library was run on a 1.8% agarose gel for 45 min at 120 V to obtain clear demarcated bands. The band was purified twice using QIAquick Gel Extraction Kit and Agencourt AMPure XP magnetic beads (Beckman Coulter USA) and then analysed on a 1.8% agarose gel (Fig 3.4). Once the library was purified, pooled concentration of *STAG2* was determined and was found to be 16 ng/µl.

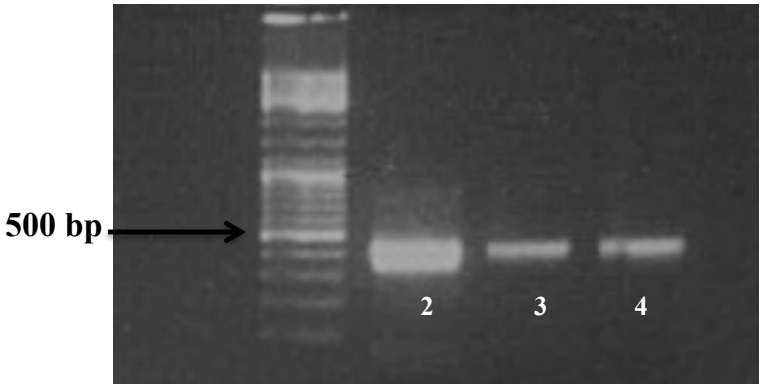


Figure 3.4: Analysis of pooled *STAG2* library on agarose gel after gel and bead purification. Lane 1: 1-10000kb ladder, Lane 2: Library after gel purification, Lane 3-4: Library after bead purification.

3.3 Sequencing of the Library and identification of Mutations

The library was sequenced on the Roche 454 GS FLX sequencer and the data was analysed using Roche Variant Analysis Software. Somatic mutations were identified in six patients (4%) with an average mutant allele burden of >15%. The mutant allele burden is defined as the ratio between reads contained in the mutant allele vs wildtype alleles, which are summarized in (Table 3.3). Average sequencing coverage depth (total number of nucleotides from reads that are mapped to a given position) across the whole genome was greater than 400X (Fig 3.5 and 3.6). All the mutations were confirmed by Sanger Sequencing on constitutional DNA except for one patient for which no constitutional source was available. Somatic nature of the mutation for this patient was confirmed by the absence of mutation in any data base i.e. dbSNP or 1000 genome project that ruled out polymorphism.

Table 3.3: *STAG2* mutations type allele burden and sites. Distribution of *STAG2* mutations found in six MDS patients. The mutations were identified in different exons and were predominantly nonsense mutations with an average mutant allele burden of >15%.

Patients	Exons	Nucleotide change	Amino Acid Change	Mutant allele burden (MAB)
1.	6	CGA>TGA	R110X	32.98%
2.	19	TTG>TAG	L589X	50.10%
3.	20	del TC_ins GGG	F647fsX	21.05%
4.	26	CAG>TAG	Q888X	34.26%
5.	27	AGG>AGC	R908S	18.27%
6.	28	CGA>TGA	R953X	43.16%

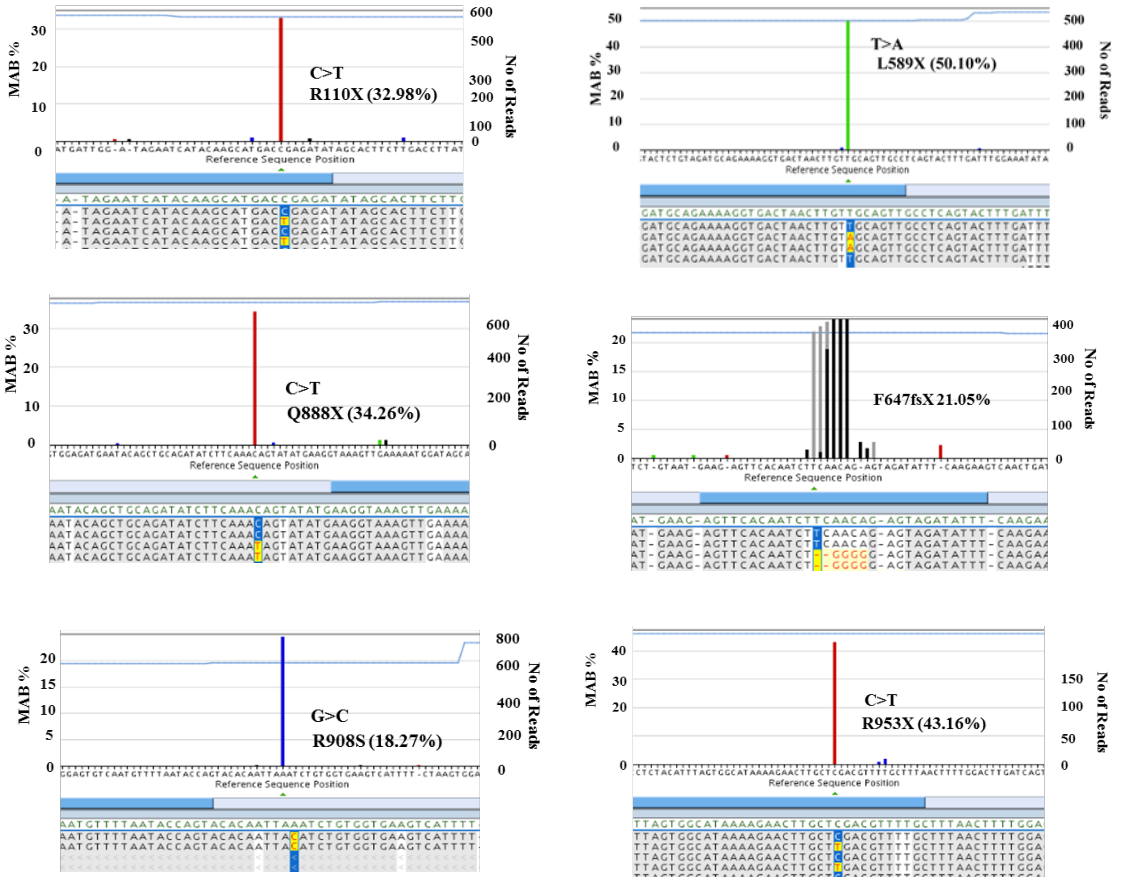


Figure 3.5: Distribution of generated amplicon sequence reads per patient. For all the patients, the read length (x- axis) and number of generated sequence reads (y-axis) is represented. Mutation allele burden is on the (y axis).

3.3.1 Mapping of mutations to different regions of STAG2 protein

Mutations were located in exons 6, 19, 20, 26, 27, 28. All the mutations were heterozygous. Four of the mutations were nonsense mutations, one was a frameshift deletion and other was a missense mutation. All the mutations were outside the two major SCD (Stromalin conservative domain) and STAG domain found in STAG-2 protein (Fig 3.6). STAG2 has been associated with gene regulation through the formation of loops which involve CTCF (CCTC binding factor). Interaction between CTCF and STAG2 has been reported to occur in the region between 162 to 993 amino acids (Xiao et al., 2011). Five of the mutations were identified in this region.

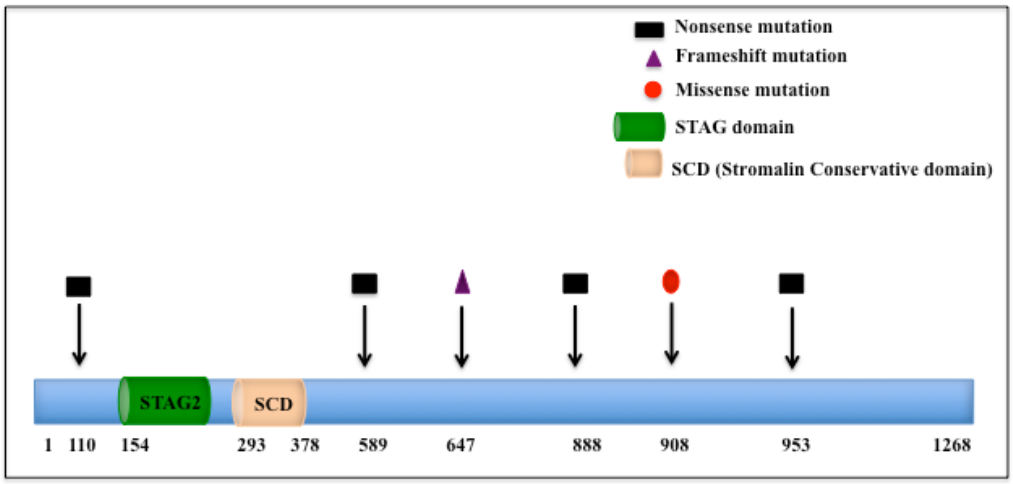


Figure 3.6: Somatic *STAG2* mutations in MDS patients. Distribution of mutations across STAG2 protein. STAG2 is a 1268 amino acid protein and has two main domains STAG2 and Stromalin conservative domain (SCD). Most of the mutations in *STAG2* were nonsense mutations resulting in truncation of protein and occurred outside the STAG2 and SCD domain.

To investigate the impact of mutations an Insilco 3D model of STAG2 was constructed using the Swiss model (Biasini et al., 2014). Although sequence homology predicted by software was 100% but the coverage of the protein sequence contained residues starting from 83-1207 (Fig 3.7). Crystal structure has

been predicted for yeast *Zygosaccharomyces rouxii* (*Zr Scc3*) residues 88-1035 amino acids. Scc3 is an ortholog for human STAG2. As sequence homology among Scc3 ortholog stretches throughout the entire Z.rouxii structure most if not all features are likely to be shared by orthologs from a wide variety of eukaryote including animals. The crystal structure revealed a partly twisted and partly crescent shaped structure composed of large number of α helices stacked upon one another some of which contained the signature motif Asp19/Arg25 found in HEAT repeats. Three HEAT repeat motifs were found between residues 341-450. In addition to a pronounced hook at the proteins C-terminal half, a prominent feature was a nose at the N terminal part at the tip of which three basic residues KKR (298-300) were found. None of the mutations were found in the HEAT domain N terminal part at the tip of which three basic residues KKR (298- 300) were found.

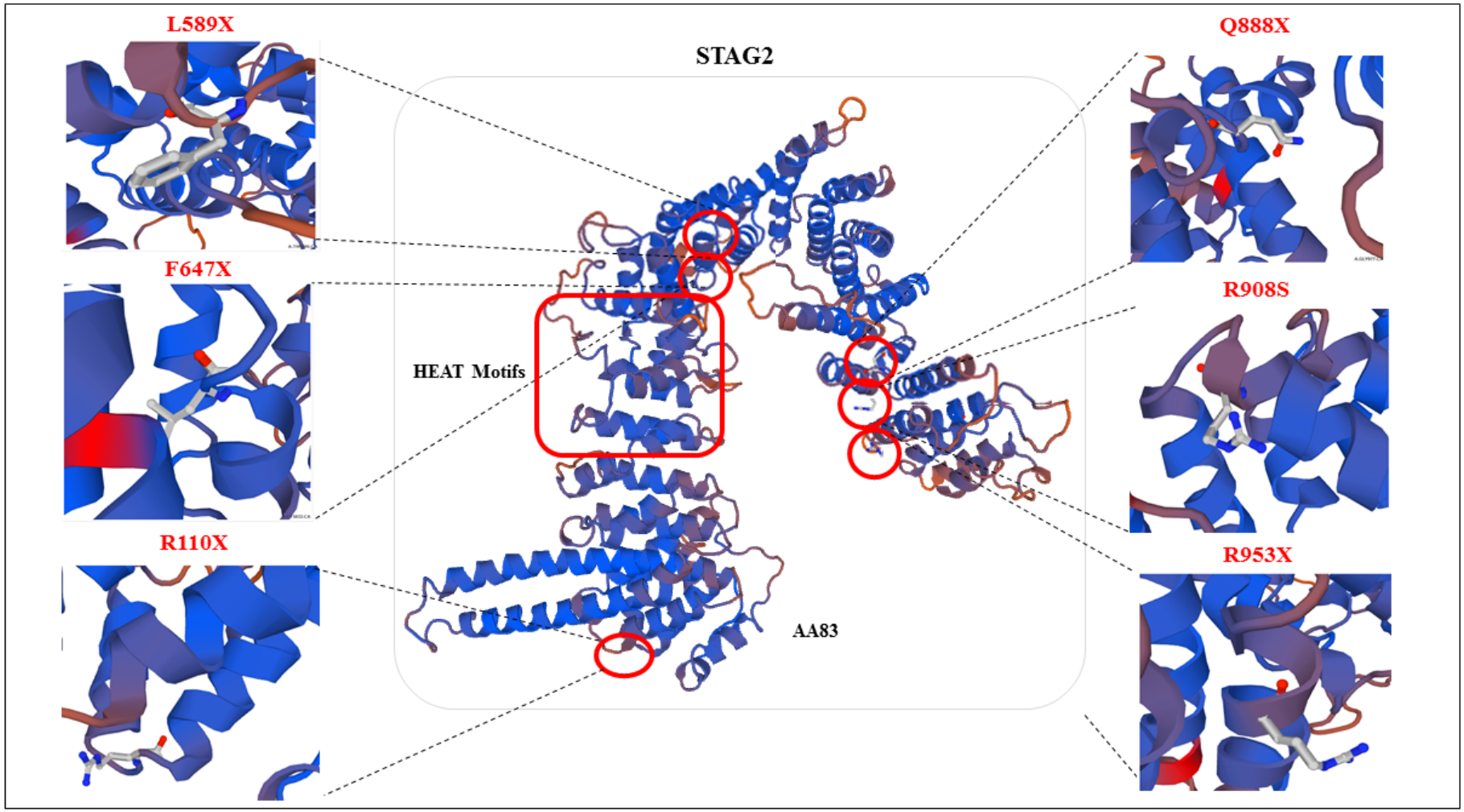


Figure 3.7: 3D structure of STAG2 protein (centre) and detailed structure of mutated codons (right and left). Mutated codons are represented within the red circle centre by small structures red and blue. The square structure represents the HEAT domains

3.4 Testing for *RAD21*, *SMC1A* and *SMC3* Genes

DNA was extracted using the Qiagen extraction kit (Refer to section 2.14.1 for protocol) for 154 patients but was amplified for 72 patients using the amplification DNA kit from Qiagen (Refer to section 2.14.3 for more details) due to low cell number.

3.4.1 Primer testing and PCR Optimisation

All primers were designed using the Primer 3 sofware http://bioinfo.ut.ee/primer3/ and tested on 50 ng/µl of genomic DNA (Fig 3.8). While testing for primer efficiency, an extra band was observed in amplicon 5 for *SMC1A* gene and thus a gradient PCR for amplicon 5 was performed at three different annealing temperatures 62°C, 64°C and 66°C (Fig 3.9). No extra band was detected at these temperatures and thus PCR for all the amplicons for all the three genes was performed at annealing temperature 62°C apart from amplicon 3 for SMC3 which was performed at 60°C as it didn't work at above three annealing temperatures. All PCR reactions were optimized using 4ng/µl of patient DNA and 5µM of primer (Refer to section 2.15.1 for more details). Once PCR was performed for all the three genes, the same four patients for each gene were run on a 1.5% agarose gel at 120V. After adjusting the concentration for all the amplicons for each patient for all the three genes, all the amplicons for each patient for each gene were pooled together. Gel and bead purification was performed using QIAquick Gel Extraction Kit and Agencourt AMPure XP magnetic beads from Beckman Coulter (USA) to purify the PCR mixes (Fig 3.10). Two picogreen measurement were performed and then volume was calculated to get a concentration of 0.2ng/µl to be used for Nextera PCR.

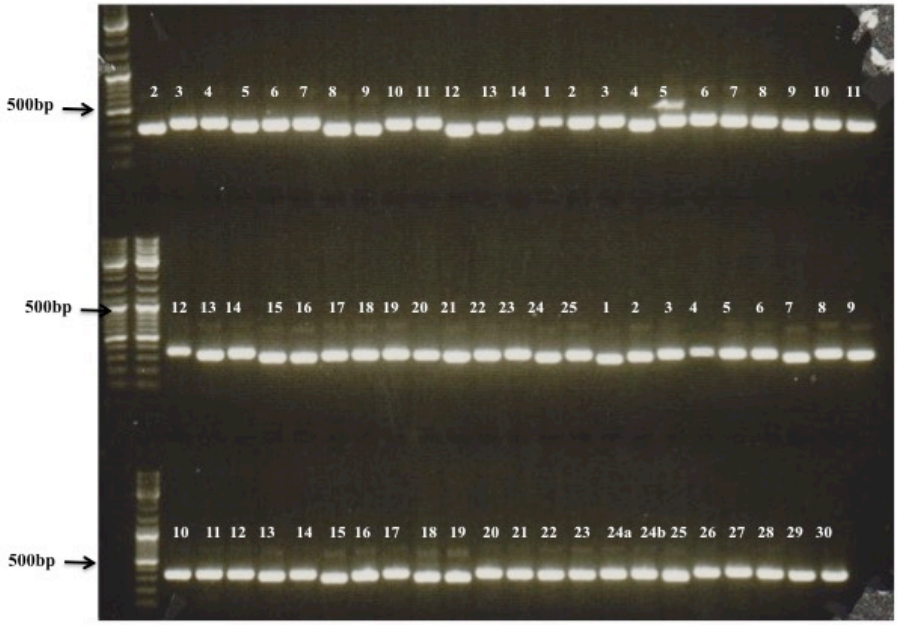


Figure 3.8: Testing for Primer efficiency for *RAD21*, *SMC1A*, and *SMC3*. All the amplicons for the three genes were tested on non amplified DNA. Upper PCR panel: 2-14 amplicons for *RAD21,* 1-11 amplicons for *SMC1A*. Middle PCR panel: 12-25 amplicons for *SMC1A*, 1-9 amplicons for *SMC3*. Lower PCR panel: 10-30 amplicons for *SMC3*.

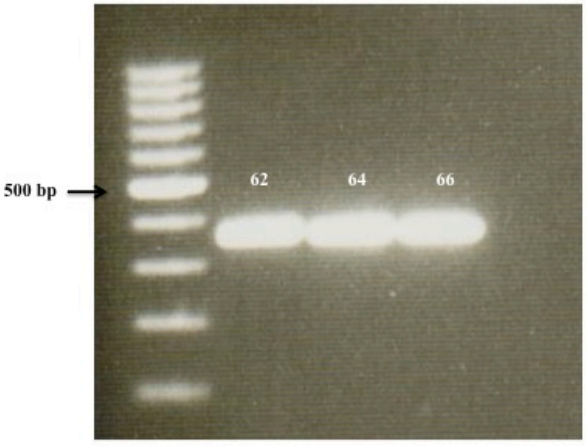


Figure 3.9: Gradient PCR at three different annealing temperatures for amplicon 5 of *SMC1A*. Gradient PCR worked at all the three different annealing temperatures 62°C, 64°C and 65°C for amplicon 5 of *SMC1A* so PCR was then performed at 62°C.

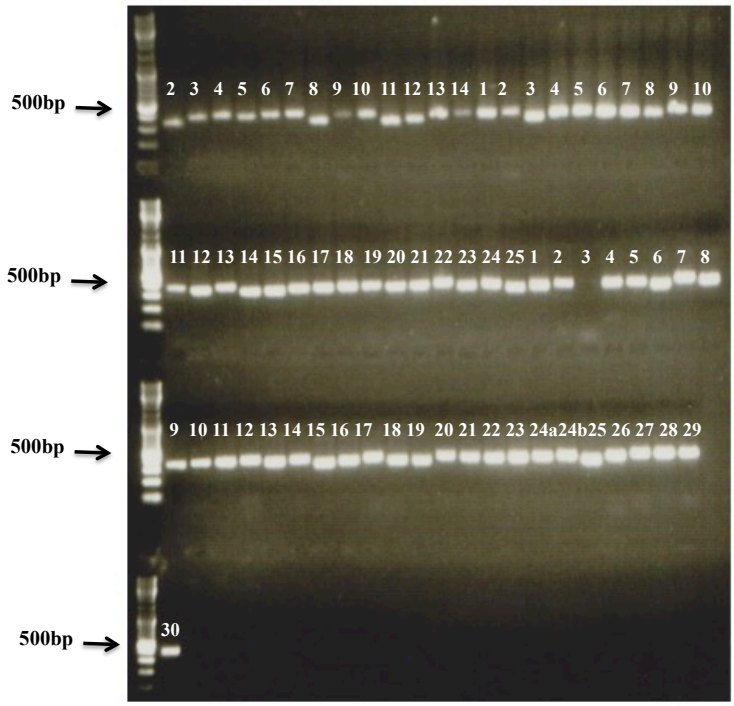


Figure 3.10: Visualization of all the amplicons for all the three genes from one patient sample after 1st round of PCR. Upper PCR panel 2-14 for *RAD21* 1-10 for *SMC1A*. Middle PCR panel 11-25 for *SMC1A*, 1-8 for SMC3. Lower bottom two PCR panel for *SMC3.*

3.4.2 Nextera PCR

For the Nextera PCR, the input DNA is tagmented (simultaneously tagged and fragmented) by the Nextera XT transposome technology. The Nextera XT transposome simultaneously fragments the input DNA and adds adaptor sequences to the ends allowing amplification by PCR in subsequent steps (Refer to section 2.20.5 for more details). After the PCR was performed, the plate was kept at -20°C for two days, followed by bead purification. Nextera libraries were quantified using picogreen assay. Calculations was made in order to achieve a final concentration of 4nM DNA followed by pooling together of all the patient samples and performing another picogreen assay.

3.4.3 MiSeq Amplicon Sequencing and Analysis

The MiSeq sequencing technology was used to sequence the pooled library which was denatured and diluted before being loaded on the flow cell for sequencing (Refer to section 20.18.2 for more details) PhiX control was used as an internal control. The data was analysed using the variant studio. Since no constitutional source was available the mutations were confirmed using independent PCR and also checked in various data bases i.e. dbSNP and 1000 genome project to rule out for polymorphism.

3.5 Identifications of Mutations

Mutations were identified in 3 out of 154 patients (2%). A missense mutation in *SMC1A* was found in one patient and a frameshift insertion in one patient and an intronic mutation in 3' splice site of exon 22 in another patients for *SMC3* (Table 3.4). Mutation was not detected in *RAD21* gene. Mutations were detected in the coiled coiled domain of both *SMC1A* and *SMC3* protein. Mutations in the intermolecular domains have been reported to either disrupt binding of accessory proteins to cohesin ring or disrupt the inter or intramolecular interaction of head domains (Fig 3.11).

Table 3.4 Patients with identified *SMC1Aand SMC3* mutations along with the variant allele frequency. Mutations were identified in three patients, one in SMC1A and two in *SMC3* gene. All the mutations were exclusive of one another.

Patient	Gene	Exon	Nucleotide change	Amino acid Change	Variant Allele Frequency
1.	*SMC1A*	13	CGG>TGG	R>W	12.93%
2.	*SMC3*	22 (Intronic)	A>G	SA 3'Exon 22	30.08%
3.	*SMC3*	24	AAA_GAA ins AAA	K939_E937 insK	9.4%

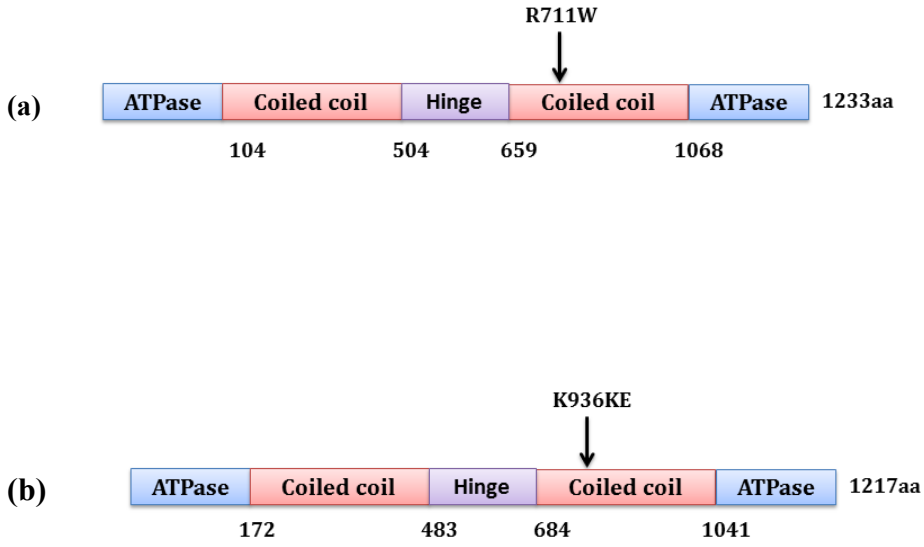


Figure 3.11: Mapping of the mutations to various domains of SMC1A and SMC3 protein.(a) SMC1A and (b) SMC3 are members of the structural maintenance of chromosome proteins. SMC1A and SMC3 proteins are 1233 and 1217 amino acids long with a molecular weight of 143 and 141 kDa respectively. They have identical structure and contain five domains, two ATPase and coiled coil domains and one hinge domain.

Second mutation in *SMC3* has not been depicted as it was found in the intronic sequence. An Insilico 3D model could not be constructed for SMC1A and SMC3 proteins using the Swiss model software as no homology sequence was detected by the software.

3.5.1 Correlation between cohesin complex mutations and clinical phenotype.

Sequencing of the Cohesin complex revealed mutations in 6% of patients (9/154) comprising *STAG2* 4% (6/154), *SMC1A* and *SMC3* 2% (3/154). WHO subgroups for patients with cohesin mutations included RAEB1 (3/154, 2%), RAEB2 (2/154, 1.28%), RCMD (2/154, 1.28%), MDS/MPN (1/154) and sAML (1/154) (Table 3.5). According the IPSS-R (International Prognostic Scoring System- Revised) cohesin mutations were more common in Very High (3/9 33.3%) as compared to Very Low (1/9 11.1%) category. Mutations were identified in six males and three females patients. The median age of the patients at the time of diagnosis was 74 years. Five

out of six aberrations detected in *STAG2* gene were nonsense mutations resulting in truncation of the protein, a missense mutation was found in *SMC1A* and a frameshift insertion and a mutation in intronic splice site in *SMC3*. Six of the patients showed progression to AML. Four males and one female showed normal karyotype by metaphase cytogenetics which was performed at Kings College Hospital. Mutations in cohesin complex genes were correlated with coexisting mutations in genes involved in various pathways done as a part of study in the same cohort of patients using a 22-gene panel by *Mian et al* (Mian et al., 2013). Coexisting mutations were found in genes involved in epigenetic, splicing factor, cell signalling pathway and transcriptional regulation (Table 3.5). Mutations in gene encoding epigenetic factor *ASXL1* (Additional sex comb like protein 1) and splicing factor *SRSF2* (Serine arginine splicing factor 2) were the most common aberrations found in patients with cohesin patients. All the mutations identified in *ASXL1* were nonsense mutations resulting in truncation of protein while hotspot mutation (P95) was found in most of the *SRSF2* mutated patients. In three of the patients *ASXL1* mutations co-existed with *SRSF2* mutation. WHO subgroups for patients with *ASXL1* included three with RAEB1 and one with RAEB2 while for *SRSF2* two RAEB1 and one each of RAEB2 and RCMD patients were found. Further according to the IPSS-R scoring system only one of the three patients with cohesin mutations found in Very High category had a coexisting mutation in *SRSF2* and *ASXL1*.

All the patients with cohesin mutations showed a shorter progression free survival as compared to cohesin wildtype patients (p=. 001) with a median free survival of 8.5 months. Median overall survival of 32 months with cohesin mutations as compared to 33 months without mutations (p=0.86). (Fig 3.13)

Table 3.5 Clinical characteristics of all the patients with cohesin mutations.

Gene	Mutations	Age	Gender	Blast	Diagnosis	IPSS-R	Metaphase Cytogenetics	Variant allele frequency (VAF)	Splicesite Mutation	Epigenetic Mutation	Cell signaling pathway, Transcriptional Regulator and other gene mutations	Progression to AML
STAG2	R110X	69	M	2	RCMD	Very Low	46,XY	32.98%		*TET2* C1193W 49% *EZH2* R690H 70%		No
STAG2	R908S	75	F	23	sAML	Very High	trisomy6 and 21 del11q(q25:q21)	18.27%				
STAG2	L589X	69	M	9	RAEB1	Intermediate	46,XY	50.10%	*SRSF2* P95L 41%	*ASXL1* Q760X 37%	*RUNX1 SA 3'Exon 4 32%*	Yes
STAG2	F647fsX	79	F	19	RAEB2	Very High	46,XX,t(7;21),der19	21.05%	*SRSF2* P95L 45%	*ASXL1* R693X 26%	*FLT3-ITD* R595-dup5%, *NRAS* Q61R 11%	Yes
STAG2	R953X	67	M	13	RAEB2	Very High	46,XY	43.16%				
STAG2	Q888X	82	F	9	RAEB1	Low	46,XX	34.26%	*SRSF2* P95L 50%	*ASXL1* G646WfsX12 22%		Yes
SMC1A	R711W	79	M	6	RAEB1	High	45,XY-7[3]46XY[5]	12.93%		*ASXL1* G646WfsX12 15%	*NRAS G12V* 23%	No
SMC3	Splice site	73	M	0	RCMD	Low	trisomy19, trisomy8	30.08%	*SRSF2* Y93fsX121 24%			Yes
SMC3	K936-E93 insK	69	M	4	MDS MPN	N/A White count High	46,XY	9.4%		*EZH2* SA 5'Exon 19 G>C 20% *TET2* C1273F 71%		Yes

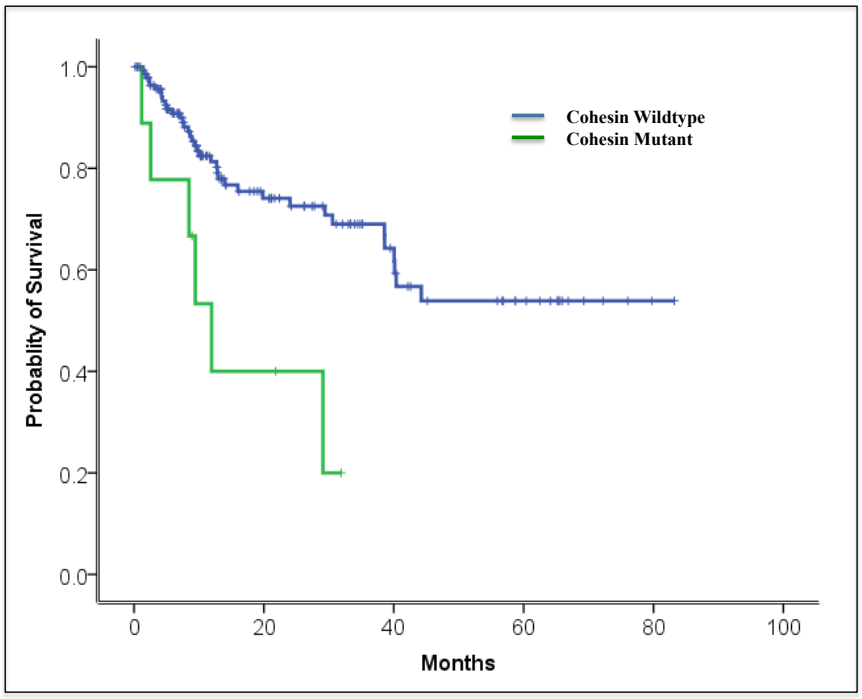


Figure 3.12: Progression free survival in 154 patients with MDS/secondary AML. Kaplan - Meier plot for progression free survival for patients with cohesin mutation (n=9) as compared to wildtype patients (n=145) p = .001. Median progression free survival 8.5 months versus non –remission (NR).

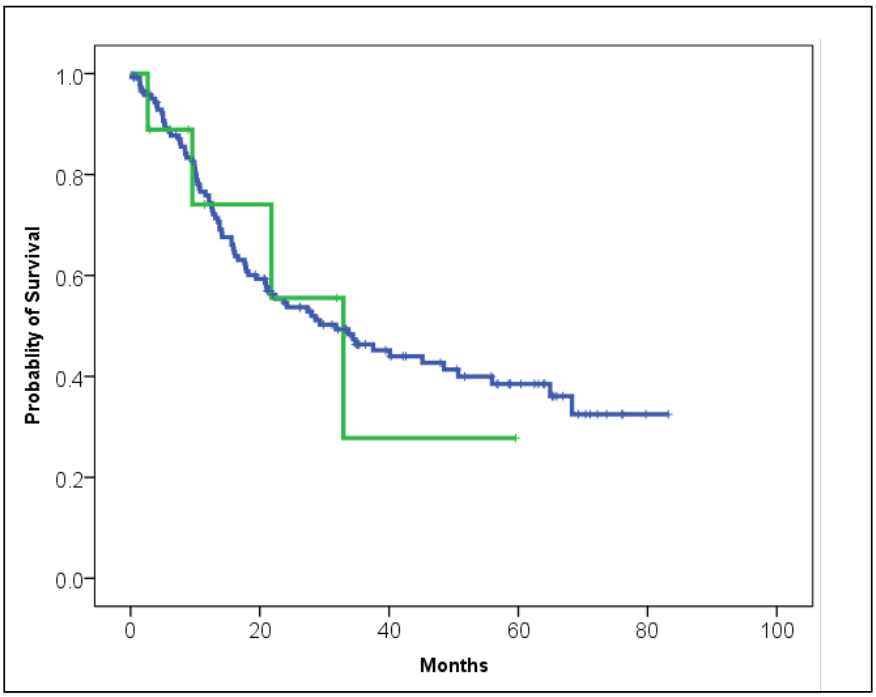


Figure 3.13 Overall survival in 154 patients with MDS/secondary AML. Kaplan - Meier plot for overall survival for patients with cohesin mutation (n=9) as compared to wildtype patients (n=145). Median overall survival 32 months vs 33 months p= 0.86. Median overall survival 34 months vs 32 months if the data is censored for haemopoietic stem cell transplant p=0.71.

3.6 Discussion

A broad spectrum of diseases arise due to germline mutations in cohesin complex genes and are termed as cohesinopathies which include Cornelia de Lange Syndrome (CdLS), Roberts syndrome/SC- Phocomelia and Warsaw breakage syndrome. Patients with germline cohesin mutations show a constellation of phenotypes including craniofacial, heart, gastrointestinal defects, poor growth, developmental delay and intellectual disability.

We identified a total of 9 patients (6%) with somatic mutations in one of the cohesin genes. Six of the mutations were in *STAG2*, one in *SMC1A* and two in *SMC3*. No aberrations were found in *RAD21*. All the mutations in our study were mutually exclusive of one another. Four of the patients with cohesin mutations progressed to AML. Patients with cohesin mutations also showed a significantly shorter progression free survival (p= .001) as compared to those without mutations. According to IPPSS-R scoring system three of the patients with cohesin mutations were found in very high subgroup. Cohesin mutations fall into two categories, truncation and frameshift mutations reported in *STAG2* and *RAD21* and missense mutations reported in *SMC1A* and *SMC3* (Kon et al., 2013).

STAG2 is located on chromosome X q25 and is mutated in 6/154 (4%) patients in our cohort, four males and two females. Four of these mutations were nonsense mutations suggestive of C terminal truncation of the protein, the other two were missense and frameshift mutation. One of the nonsense mutation arginine to a stop codon (R110X) has been reported in two different studies. *Kon et al* reported this mutation in two patients with Chronic Myeloid Leukaemia in blast crisis (CML) and Chronic Myelomonocytic Leukaemia (CMML) (Kon et al., 2013) while *Thota et al* reported this mutation in patient with Refractory cytopenia with multilineage dysplasia (RCMD) (Thota et al., 2014). In both the studies, the reported patients

were males and showed normal karyotype, except in the CMML patient where the karyotype was undetermined. Although the position of the reported mutation was different in all the patients, the mutation would have resulted in premature truncation and reduced expression of protein. All the *STAG2* mutations identified in our study were located outside the two main domains found in STAG2 protein, the SCD (Stromalin Conservative Domain) and STAG2 domain. Five of the mutations were located in the region between (162-993) amino acids which could have affected interaction of *STAG2* with CTCF. STAG2 has been reported to interact with CTCF in this region (Xiao et al., 2011). CTCF a 11 zinc finger DNA binding protein is the only insulator protein reported in vertebrates which can block enhancer function in plasmids and native DNA context. Interaction of STAG2 with CTCF is reported to be crucial for the expression of imprinted *Igf2*/*H19* genes and insulation activity of CTCF. Also direct contact between CTCF and cohesin has been found to occur via STAG2. Phosphorylation sites have been identified in STAG2 at residues S1058, S1064, S1065 in the C terminal region that is important for the dissociation of cohesin from chromosome arm during mitosis. Phosphorylation of *STAG2* might also be important for recruiting cohesin unloading factors to chromatin that can then enable dissociation of cohesin from chromosomes (Hauf et al., 2005). As truncated proteins would have been the end product of all the non-sense mutations in our study this could have an impact on sister chromatid separation due to loss of phosphorylation sites.

SMC1A gene is located on chromosome Xp11.2. One patient in our cohort had a missense mutation (R711W) in *SMC1A* gene. The patient was a male. This mutation which is associated with a Cornelia de Lange phenotype (CdLs) was found to be somatic residing in the coiled coil domain. By use of the Coils program

which predicts the probability of a protein to form a coiled-coil *Deadorff et al* reported that this mutation had a small likelihood of disrupting the coiled-coil arms. However, the alterations caused by this mutation might affect the angulation of the coiled-coil resulting in impaired intra or intermolecular approximation of the SMC head domains or disrupt binding of accessory proteins to the cohesin ring (Deardorff et al., 2007).

SMC3 gene is localised on chromosome 10 q25.2. Two patients in our cohort had mutations in *SMC3* gene, one was a frameshift insertion in the coiled coiled domain of exon 24 and the other was in the 3'splice intronic acceptor site of exon 22. Studies have shown that mutations occurring in the bases at the beginning of the exon-intron junction have an impact on the splicing of the intron and result in the production of defective protein (Raben et al., 1992). Both the patients were males and showed progression to AML.

All the truncating mutations in cohesin complex were identified in *STAG2* gene. Solomon *et al* identified a diverse range of tumors harbouring deletions or inactivating mutations of *STAG2* and correlated aneuploidy with mutational inactivation of *STAG2*. As the presence of p53 mutations is characterised by gross chromosomal instability, it is interesting that none of the cohesin mutated patients had a p53 mutation. A number of papers have documented mutations in cohesin complex are not to be associated with chromosomal instability (Taylor et al., 2013, Welch et al., 2012) These result suggest that either loss of cohesin is not associated with chromosomal instability in the background of wildtype p53 or there might be a compensatory mechanism that compensates for the loss of cohesin complex genes and hence survival of cohesin deficient tumour cells. *STAG1* is a paralog of *STAG2* primarily involved in telomere replication. *Solomon et al* and other groups could not

detect an upregulation of STAG1 expression in STAG2 deficient cells. In contrast, *Kon et al* reported compensatory upregulation of STAG1 or STAG2 in SA1/SA2 deficient HeLa cells. *Bailey et al* also showed an increase in STAG1 containing complexes in *STAG2* mutated glioblastoma cell line following immunoprecipitation with RAD21 (Bailey et al., 2013). Interestingly a recent study carried out by *Kim et al* showed that chromosomal instability is not the phenotype of all cohesin mutations and different mutations have either different mechanisms by which they cause cancer or the mechanism is unknown (Kim et al., 2016). In this study all the missense mutations tested retained wildtype cohesion, mitotic integrity and euploidy also the truncating mutations which despite showing defects in sister chromatid cohesion and a subset showing increase in lagging chromosome didn't show any change in modal chromosome number. Further they demonstrated that truncating mutations in *STAG2* had no influence on its interaction with other members of the cohesin complex and supported this hypothesis with the recently published crystal structure of STAG2 which showed that missense mutations didn't have any influence on interaction of STAG2 with members of the cohesin complex.

as STAG2 interacts with cohesin via an extensive interface with RAD21 spanning the entire length of *S*TAG2 (Hara et al., 2014). Further correction of endogenous *STAG2* mutant allele in two isogenic glioblastoma cell lines by somatic cell gene targeting did not disrupt the ability of SMC3 to co-immunoprecipitate SMC1A, SMC3 and RAD21 subunits suggesting that STAG2 is not required for assembly of cohesin subunits .The results from this study could support our findings that despite truncating mutations in *STAG2* no influence on chromosomal instability was observed as STAG2 is not responsible for formation of the ring.

In five of our patients, cohesin mutations were found to co-occur with mutations in *SRSF2* (ribosome pre-mRNA splicing factor) and *ASXL1* (polycomb repressive transcriptional factor mutations) genes. *SRSF2* (Serine/ arginine rich splicing factor 2) is a member of serine /arginine- rich (SR) protein family that contributes to both constitutive and alternate splicing by binding to exonic splicing enhancer (ESE) sequences within the pre-mRNA through its RNA recognition motif domain (RRM). Most of the splicesome mutations are heterozygous, mutually exclusive and are associated with adverse outcomes among MDS and acute myeloid leukemia (AML) patients. *Kim et al* associated hot spot mutations (P95H/L/R) in *SRSF2* to be responsible for alterations in splicing of major haematopoietic transcriptional regulators like *EZH2*. A change in the conformation of the termini of *SRSF2* RRM caused retention of poison exon cassette in *EZH2* resulting in nonsense mediated decay (NMD) of *EZH2.* Mutations in *SRSF2* were identified in four of the cohesin mutated patients with three patients carrying hot spot (P95H and P95L) mutations in *SRSF2*. In most of the cases in our cohort aberrations in *SRSF2* coexisted with nonsense mutations in cohesin complex. Impaired *SRSF2* could have resulted in altered splicing of cohesin subunits subsequently leading to nonsense mediated decay however this effect could not be ascertained at a protein level due to unavailability of cells from these patients (Kim et al., 2015). This hypothesis can further be supported by a study carried out by *Sundermoorthy et al* where using a functional genomic approach an association between pre mRNA splicing factors and sister chromatid cohesion was unveiled. Loss of a multitude of splicing factors including *SF3B1* (one of the genes of the splicesome machinery) was found to cause a loss of sister chromatid cohesion upon mitotic entry. Compromised splicing produced a non functional truncated sororin (a protein required for stable

association of cohesin with chromatin) due to retention of intron 1 in sororins mRNA. Reduced levels of sororin were detected in cells depleted of splicing factors. Expression of an intronless version of sororin and depletion of *WAPL* whose activity is antagonized by sororin was able to restore sister chromatid cohesion (Sundaramoorthy et al., 2014). Reported findings from the two studies predict that mutations in components of splicing machinery can result in compromised splicing of cohesin members which could alter the turnover of cohesin on chromatin hence affecting chromosomal stability and gene regulation.

Mutations in additional sex -combs like-1' *ASXL1* gene have been also found in cohesin mutated patients. *ASXL1* encodes a chromatin binding protein involved in epigenetic regulation and is responsible for recruiting polycomb and trithorax complexes to specific loci. Mutations in cohesin complex have been reported as coexisting passenger mutations along with *ASXL1* mutations that exist as drivers of clonal evolution (Thota et al., 2014). Clonal dominance at the time of transformation to more aggressive disease and strong association with *ASXL1* indicated synergism between epigenetic dysregulation and mechanism of action exerted by cohesin haploinsufficiency. Further mutations in *ASXL1* were thought to promote myeloid transformation through loss of PRC2 mediated myeloid repression suggesting that *ASXL1* might act as a scaffold for recruitment of PRC2 complex to specific loci in haematopoietic cells. Results emerging from gene expression and chromatin state data from this study identified *HOXA* cluster including *HOXA9* to be altered in the setting of *ASXL1* mutation (Abdel-Wahab et al., 2012, Fisher et al., 2016). Similar findings have also been reported with regards to mutations in cohesin complex by other studies. *Fisher et al* supported a model in which RAD21 was found to be responsible for recruitment of PRC2 to its targets i.e. HOXA7 and

HOXA9. Loss of *RAD21* increased haematopoietic stem cell renewal due to deficiency of H3K27me3 at the promoters of HOXA7 and HOXA9. As MDS is a myeloid malignancy in which multiple mutations co-operate in the progression of disease, association between mutations in *ASXL1* and cohesin complex by having an additive effect on altering the expression of Hox genes could lead to disease progression.

In summary mutations in cohesin complex have been identified in MDS and AML but how these mutations contribute to disease progression needs to be unravelled. What is the nature of these mutations whether they exist as driver or passenger mutations, how these mutations provide a growth advantage to the cancer cell needs to be determined and further can these mutations be exploited therapeutically to be used in a clinical setting.

In order to explore contributions of cohesin complex mutations in myeloid malignancies my next part of thesis would be to determine the functional consequences of knockdown of *STAG2* gene in a myeloid setting.

Chapter Four

FUNCTIONAL CONSEQUENCES OF KNOCKDOWN OF *STAG2* IN MYELOID CELL LINE

4.1 RNA Interference

The advent of RNA Interference (RNAi) technology has become a powerful tool to gain insight into critical biological processes like gene regulation and to identify novel therapeutic targets, (Bantounas et al., 2004, Leung and Whittaker, 2005, Sledz et al., 2003) (Fig 4.1).

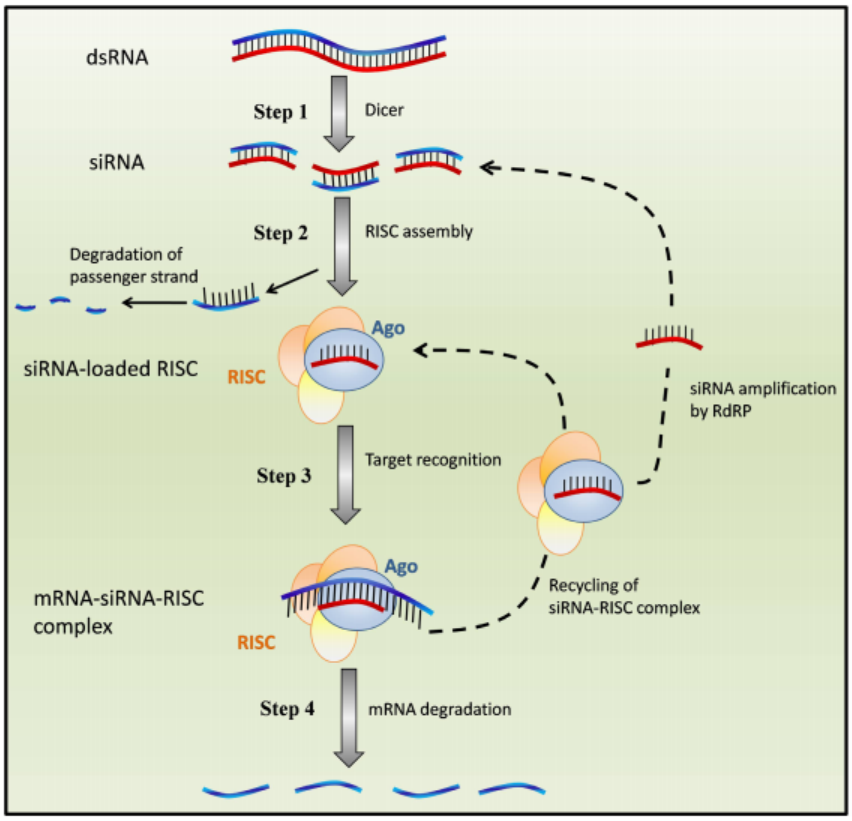


Figure 4.1: Schematic representation of RNA interference in a mammalian cell.1. The dsRNA is initially recognized by an RNase III nuclease enzyme Dicer and is cleaved into 21-23 base pairs small double stranded molecules termed siRNA. 2. siRNAs are loaded onto a multiprotein complex with RNase activity called RNA Induced Silencing Complex leading to degradation of passenger strand while the guide strand remains attached to RISC as a template in the silencing reaction. 3. The guide strand assembles into a functional siRNA-RISC and attaches to the complimentary strand. 4. The target mRNA is degraded and dissociated from the complex and the siRNA-RISC is released (Cuccato et al., 2011).

One of the drawbacks of using siRNA is the generation of off target effects that can be elicited through different mechanisms (Birmingham et al., 2006, Jackson et al., 2003, Jackson et al., 2006, Lin et al., 2005). An elicitation of an anti-viral interferon response (IR) through introduction of long dsRNA leads to widespread consequences on cellular processes such as protein synthesis, cell cycle and apoptosis. The use of synthetic siRNAs 21-23 nucleotides in length however is thought to resolve this problem (Bantounas et al., 2004). Off target effects can fall into two categories, they can be sequence dependent specific to siRNAs or sequence independent and caused by siRNA in general. Some of the off target effects can be avoided by reducing the concentration of siRNA, however most of the off targets effects are sequence dependent (Jackson and Linsley, 2004, Lin et al., 2005) Complimentarity to a 7 nucleotide region at 5' end of siRNA (between position 2-8) is not only important but sufficient enough to elicit an RNAi response (Birmingham et al., 2006, Jackson et al., 2006, Lin et al., 2005). Off target effects however do not always cause transcript cleavage but may inhibit translation as a method of suppression instead, behaving much like miRNAs (Zeng et al., 2003). Unavoidable sequence homology to a 7nt siRNA seed region else were in the genome is unavoidable making sequence specific off targets effects almost inevitable and difficult to predict. However, these off target effects are generally less efficient and much weaker by comparison to the targeted effect on the intended transcript.

A stable system of gene suppression is required to achieve long term knockdown. that can be achieved by adapting the more sustainable short hairpin (shRNA) mediated approach. Like siRNAs, shRNAs may be transfected as plasmid vectors encoding shRNAs transcribed by RNA pol III or modified pol II promoters, but can also be delivered into mammalian cells through infection of the cell with virally

produced shRNA. shRNA expression cassette consists of a sense and antisense strand separated by a loop. The RNA transcript folds at the loop and self anneals to form a dsRNA stem loop structure: a short hairpin RNA. The shRNA is structurally related to endogenous miRNAs and is processed accordingly by the RNAi machinery to mediate target silencing.

Conventionally the RNA polymerase III promoters U6 and H1 are used for shRNA expression, generating transcripts lacking a long poly (A) tail that may hinder subsequent RNAi processing machinery. However, RNA polymerase II promoters such as the CMV promoter have been used. RNA interference is an effective technique to knock down gene expression in order to study protein function in a wide range of cell types as long as off target effects are circumvented. The use of mock and scrambled controls in my work help to circumvent this non-specific effect.

4.2 Aim

- Establishment of a transient or stable *STAG2* knockdown in a myeloid cell line in order to delineate the functional consequences of *STAG2* knockdown in myeloid malignancies
- Consequences of *STAG2* knockdown on the expression of other members of the cohesin complex.
- To determine whether sensitivity to poly-ADP ribose polymerase inhibitors (PARPi) is attributable to defects in the cohesin complex using the Homologous Recombination assay (HR assay). The aim is to exploit cohesin mutations therapeutically, which could aid in the development of new drugs specifically targeting cohesin mutations.
- To study the effect of PARPi on proliferation and cell cycle in shRNA lentiviral mediated silenced *STAG2* using trypan blue exclusion and flow cytometry.

4.3 Results

4.3.1 Establishment of a transient cell line

Two predesigned siRNA oligonucleotides, one from Dharmacon™ and the other from Santa Cruz™ were used to transiently silence *STAG2* expression using electroporation (Refer to section 2.10.1 for more details). A negative control (a scrambled oligonucleotide) was used to distinguish sequence specific silencing from non-specific effects. To exclude any variations caused as a result of the technique mock-transfected cells were used. Cell lines i.e. K-562, NB4, U-937, MCF-7 were tested for knockdown efficiency using different concentrations of siRNA. Samples were collected at 24, 48 and 72hr timepoints. and maximum knockdown efficiency was achieved at 48hr timepoint in U-937 cell line using 100nM of siRNA (Fig 4.2).

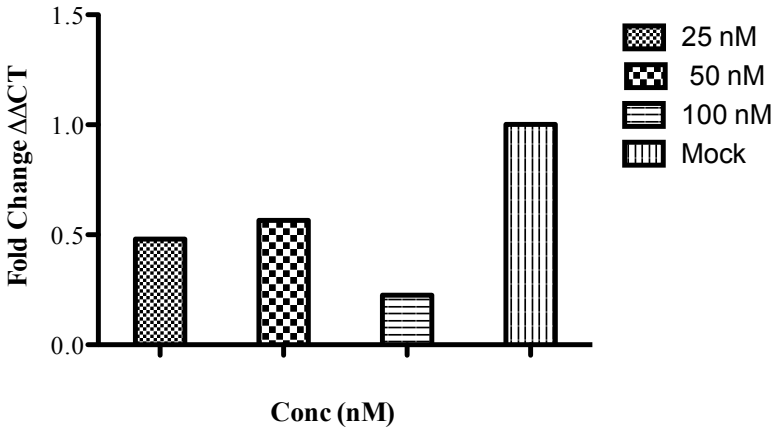


Figure 4.2: Detection of maximum *STAG2* knockdown efficiency in U-937 cell line using different concentrations of Dharmacon siRNA. In order to attain maximal knockdown efficiency of *STAG*, U-937 cells were electroporated with different concentrations of Dharmacon siRNA and maximum knockdown efficiency was achieved using a concentration of 100nM of siRNA at 48hr timepoint determined by qRT-PCR. *GAPDH* and *TUBULIN* (data not shown) were used as endogenous controls. The results are expressed as *STAG2* mRNA fold difference compared to that measured in mock treated control $\Delta\Delta C_T$.

4.3.2 Transient knockdown of *STAG2* using electroporation method

The viability and general health of cells prior to transfection is considered to be an important source of variability from one transfection to another. Electroporation a very harsh technique uses electric impulses to create temporary pores in the cell membrane through which substances like nucleic acid can pass. In preliminary experiments, a substantial amount of cell death was observed, therefore a starting cell density of 2×10^6 cells/ml was found to be efficient to achieve maximal transfection efficiency (Refer to section 2.10.1 for more details). U-937 cells were then transfected with 100nM of siRNA and samples were collected at 24, 48 and 72hr time point and knockdown was confirmed by qRT-PCR (Fig 4.3).

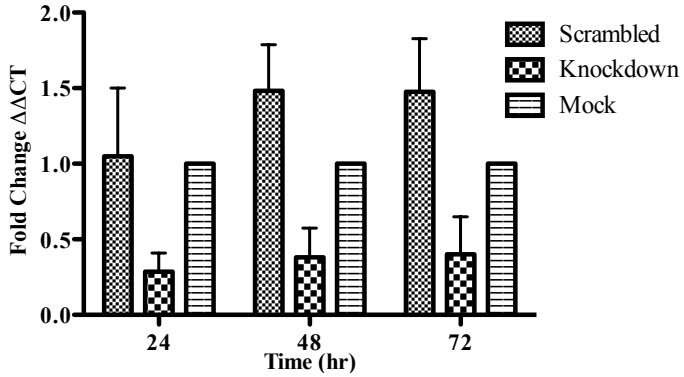


Figure 4.3: Silencing efficiency of Dharmacon siRNA in U-937 cell line. Expression level of *STAG2* (using 100nM of Dharmacon siRNA) was calculated relative to mock-infected cells (ΔΔ CT) by q-RT PCR. *GAPDH* and *TUBULIN* (data not shown) were used as endogenous controls. Maximum knockdown (62%) was achieved at 24hr timepoint with a gradual increase in expression at later timepoints. Data presented as mean ±SD (n=3).

Knockdown efficiency of around 63% was achieved that was observed at the 24hr timepoint with a gradual increase in expression over time. Therefore the 24hr time point was considered to be the optimal timepoint to carry out further experiments. Expression levels for *STAG1,* a paralog of *STAG2* was also carried out to determine whether upregulation of *STAG1* occurs upon knockdown of *STAG2* as previously reported (Bailey et al., 2013) (Fig 4.4).

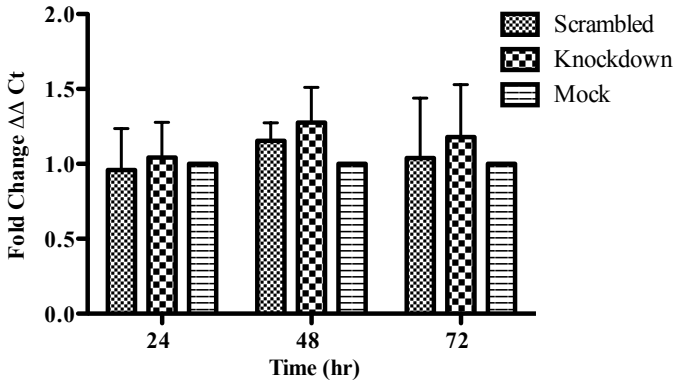


Figure 4.4: Expression levels of *STAG1* in *STAG2* si-RNA silenced (Dharmacon). U-937 cell line determined by qRT-PCR. Expression level of *STAG1* in *STAG2* knockdown U-937 cell line (using 100nM siRNA) was calculated relative to mock-infected cells (ΔΔ CT). *GAPDH* was used as an endogenous control. No upregulation of *STAG1* was observed in knockdown samples. Data presented as mean ±SD (n=3).

No significant upregulation of *STAG1* was observed in *STAG2* silenced U-937 cells at 24hr. To confirm that the knockdown by siRNA (Dharmacon) was real and not artifactual, another siRNA from Santa Cruz was used at the same concentration of 100nM using electroporation method (Fig 4.5).

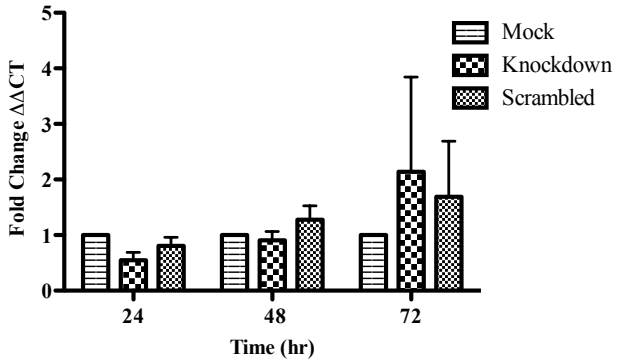


Figure 4.5: Expression level of *STAG2* in U-937 cell line transfected with siRNA (Santa Cruz) and determined by qRT-PCR. Expression of *STAG2* in U-937 using 100nM of si-RNA from Santa Cruz was calculated relative to mock-transfected cells (ΔΔ CT). *GAPDH* and *TUBULIN* (data not shown) were used as endogenous controls. Maximum knockdown was achieved at 24hr similar to Dharmacon siRNA. Data presented as mean ±SD (n=3).

Similar knockdown efficiency i.e. 63% was achieved using the Santa Cruz siRNA at 24hr timepoint confirming the results achieved using Dharmacon siRNA. The expression level of *STAG1* was also determined to confirm the results obtained by using 100nM of Dharmacon siRNA Fig (4.6).

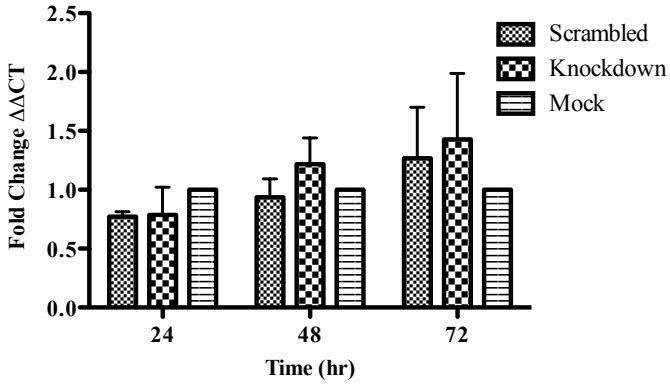


Figure 4.6: Expression levels of *STAG1* in U-937 using 100nM siRNA(Santa Cruz). Upregulation of *STAG1* was determined relative to mock-infected cells ΔΔ CT using qRT-PCR. *GAPDH* and *TUBULIN* (data not shown) were used as endogenous controls. No significant upregulation of *STAG1* was observed in *STAG2* knockdown U-937 cell line. Data presented as mean ±SD (n=3).

No effect on the levels of *STAG1* was observed with downregulation of *STAG2,* using 100nM of Santa Cruz siRNA thus confirming the Dharmacon 100nM siRNA results.

4.3.3 Cell cycle analysis

STAG2 forms a part of the cohesin complex which holds sister chromatids together and several papers have reported downregulation of *STAG2* to be associated with chromosomal instability (Solomon et al., 2011). To correlate the association between chromosomal instability and *STAG2* knockdown, U-937 cells were transfected with 100nM of Dharmacon siRNA and samples were collected at various time points up to 96 hours. Cell cycle kinetics analysis was performed using flow cytometry (Refer to section 2.5.2 for more details) to determine a change in cell cycle kinetics upon *STAG2* downregulation (Fig 4.7).

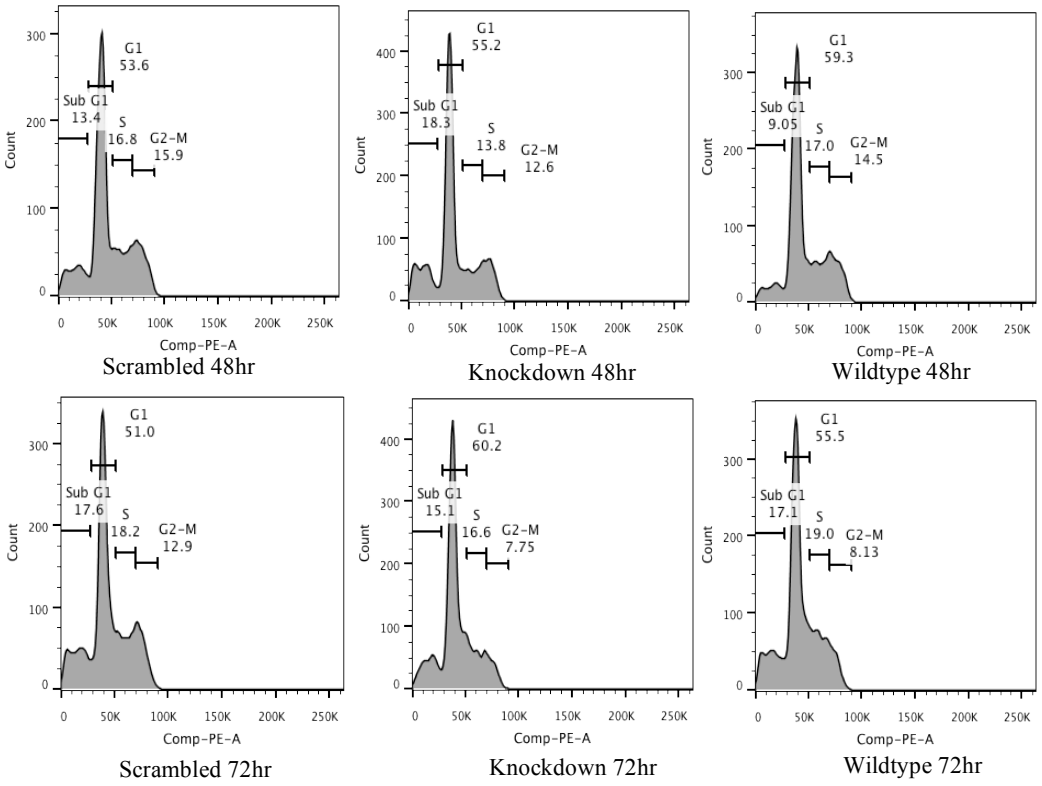


Figure 4.7: Effect of transient knockdown of *STAG2* on cell cycle kinetics. Samples were collected upto 96 hrs (data only shown for 48 hrs (upper panel) and 72 hrs (lower panel) treated with propidium iodide (for DNA content) and fluorescein isothiocyanate (for protein content) and then analyzed for cell cycle distribution using flow cytometry. The histogram was plotted using the flow jo software. The software predicted the percentage of cells in different phases of cell cycle. No significant difference was observed between the percentage of cells in different phases of cell cycle G_0, G_1, S and G2/M between the siRNA versus mock transfected cells. The result is representative of three independent experiments.

No significant differences could be determined between the scrambled, knockdown and mock transfected cells in various phases of cell cycle especially (apoptotic) G_0/sub G_1 phase. As inhibition of target by siRNA is transient, this could be one of the reasons behind the non-significant effect on cell cycle. This issue can be circumvented by the establishment of a *STAG2* downregulated stable cell line using short hairpin RNAs (shRNAs).

4.4 Stable knockdown of *STAG2* using shRNA in Retroviral system

Four independent retroviral shRNA expression vectors targeting *STAG2* were obtained from Origene (Rockville US) and were amplified by transforming *E.coli.* (Fig 4.8). Concentration of 1µg, 4µg of each individual shRNAs as well as 10 µg of pooled shRNAs were used, to transfect U-937 cells by electroporation, however cell death was observed after puromycin selection. After several unsuccessful attempts lentiviral approach was initiated to establish a stable cell line.

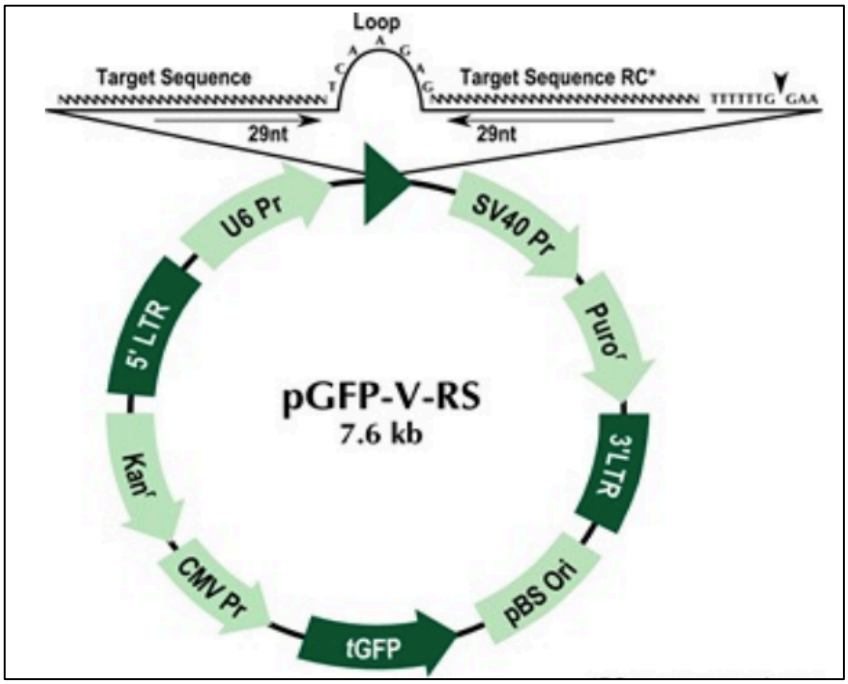


Figure 4.8: Map of shRNA cloning vector pGFP-V-RS. pGFP-V-RS is a plasmid vector used for cloning shRNA expression cassettes. The cassette consists of a 29 nucleotide target-gene-specific sequence, a 7 nt loop, and another 29 nucleotide reverse complementary sequence, all under the control of the human U6 promoter. A termination sequence (TTTTTT) is located immediately downstream of the second 29 nt reverse complementary sequence to terminate the transcription by RNA Pol III. A puromycin-N-acetyl transferase gene is located downstream of the SV40 early promoter, resulting in resistance to the antibiotic puromycin. (Image taken from origene website)

4.5 Stable knockdown of *STAG2* using a shRNA Lentiviral system

4.5.1 shRNA (Thermo Fisher)

Four independent shRNA were purchased from Thermo Fisher (Dharmacon) and virus was generated using lentivirus system (Refer to section 2.11 for more details). Cells were transduced with the virus and 48hours later transduced cells were subjected to puromycin (2µg/ml) selection. Samples were collected at day 7 and knockdown was determined using qRT-PCR, however no efficient knockdown was observed with any of the shRNA so another batch of shRNA from Addgene was obtained.

4.5.2 shRNA (Addgene)

Two shRNAs and a scrambled control were purchased from Addgene and lentiviral particles were produced (Refer to section 2.11 for more details). The shRNA vectors did not possess a fluorescent marker i.e. GFP, therefore, virus titre was roughly calculated by transfecting Hela cells with different concentration of virus and then counting the colonies. U-937 cells were transduced with 12 µl of virus for each shRNA as well as scrambled (control). Puromycin (2µg/ml) was added after 48 hrs and positive cells were collected for western and q-PCR after 14 days (Fig 4.9). 12 µl of virus was found to give an efficient knockdown of *STAG2* using both shRNAs.

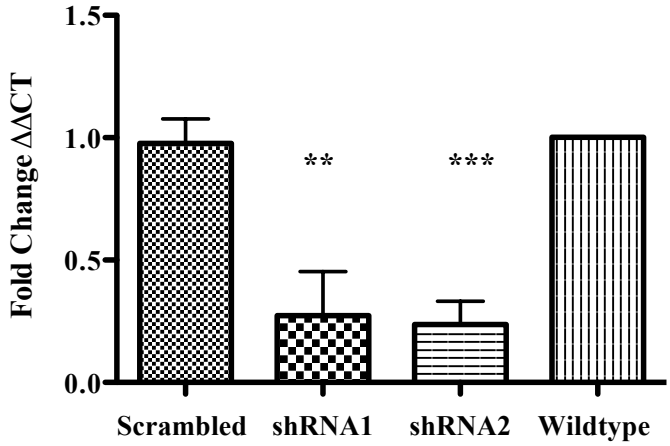


Figure 4.9: *STAG2* expression in *STAG2* lentiviral shRNA treated U-937. STAG2 expression was determined using qRT-PCR. *GAPDH* and *TUBULIN* (data not shown) were used as endogenous controls. A knockdown efficiency of 72% and 76% using shRNA1 and shRNA2 respectively was achieved. Expression levels were calculated relative to wild type *STAG2* (ΔΔ CT). Error bars represent mean of standard deviation of three independent experiments. **p<.01, ***p<.001.

A knockdown efficiency of 72% and 76% was achieved using shRNA1 and shRNA2 respectively at day 14 after post selection by puromycin. The knockdown for both shRNAs was found to be significant when compared with scrambled (control) with $p < 0.1$ for shRNA1 and $p < .01$ for shRNA2.

4.5.3 Western Blot Analysis

Reduction in STAG2 expression at the protein level was also confirmed by western blotting (Fig 4.10). To correlate the impact of *STAG2* knockdown, western blotting was also carried out on other members of the cohesin complex RAD21 (Fig 4.10 a), SMC1A and SMC3 (Fig 4.10b).

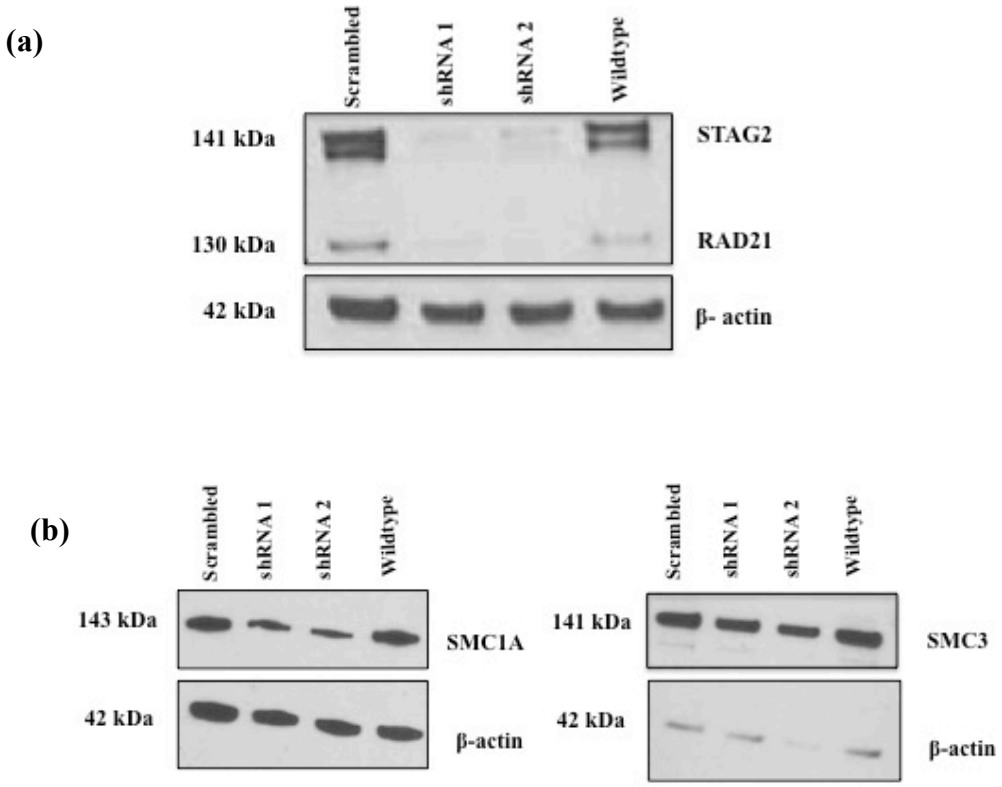


Figure 4.10: Western Blot analysis for depletion of STAG2 in lenti- virus treated U-937. Immunoblot analysis of whole cell extracts from U-937 cell line that were previously transduced with *STAG2* lentivirus (16 days prior to analysis) (a). β-actin was used as a loading control (b) Western blot was also carried out on other members of cohesin complex RAD21 (Fig 4.9a), SMC1A and SMC3 however no decrease in expression level was observed. The result is representative of five independent experiments.

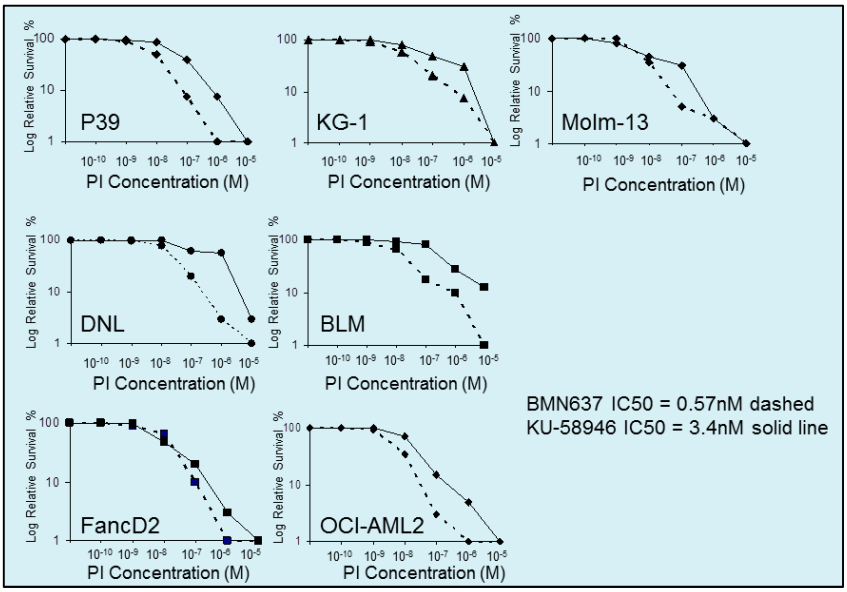
Results from western blot correlated with qRT-PCR results confirming the down regulation of STAG2 although the knockdown efficiency was more pronounced by western blot. This discrepancy in the knockdown efficiency at mRNA and protein level could not be explained. No difference in the expression levels of other members of the cohesin complex was observed although western blot did show a reduction in the expression level of *RAD21* level but the reduction could be ruled out due to unequal loading between the scrambled and knockdown samples and also the failure of confirmation at mRNA level by qRT-PCR.

4.5.4 Homologous Recombination Assay

The phenomenon of Synthetic lethality was first described by Calvin Bridges in 1992 who noticed that some combinations of mutations in the model organism *Drosophila melanogaster* confer lethality. Synthetic lethality arises when a combination of mutations in two or more genes leads to cell death, whereas a mutation in only one of these genes does not, and by itself is said to be viable.

In order to determine synthetic lethality between STAG2 and PARP. 3×10^5 U-937 cells with stable knockdown of *STAG2* were cultured in a 24 well plate in 1ml of RPMI media and 100nM of PARPi, BMN-673 (gifted by Biomarin). In a study conducted by *Gaymes et al* from our group 100nM of BMN-673 was found to be the optimal concentration to be used in HR assays as increasing the concentration resulted in non specific cell death (Fig 4.11) **(**Gaymes et al., 2013). After 24hr incubation, the cells were collected and cytospun onto poly- lysine coated slides. The cells were stained for RAD51 and γH2AXP foci (Refer to section 2.7 for more details). RAD51 foci formation is a marker for competent Homologous Recombination (HR) an error free DNA DSB repair pathway. In HR, RAD51 is recruited to sites of double stranded DNA damage where it is involved in complementary strand invasion and formation of a helical nucleoprotein filament that is detected in the form of nuclear foci. γH2AXP is a marker for DSB DNA damage. Upon phosphorylation at Ser 139, it acts as a mediator of repair factor recruitment during DSB DNA damage. The cells were visualised using a fluorescence microscope and percentage of RAD51 and γH2AX P positive cells (>5 foci) were counted (Fig 4.12, 4.13, 4.14).

(a)



(b)

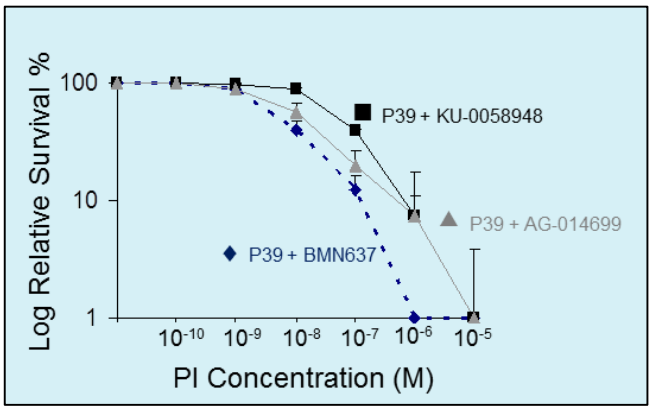


Figure 4.11: Evaluation of the effective drug concentration in leukaemic and fibroblast cell lines. (a) Soft agar clonogenic assays were used to determine cell survival in leukaemic (P-39, KG-1, Molm-13, FancD2, OCI-AML2) and fibroblastic cell lines (DNL and BLM) in response to different concentrations of BMN-673 and KU-0058948 (now known as olaparib). Data from cell survival assays were used to construct dose response curves. BMN-673 was found to be more cytotoxic than KU 0058948 at the same concentration. (b) Comparison of cytotoxicity of BMN-673 (dashed line) with KU-0058948 (solid square) and AG-014699 (solid triangle) in P-39 cell line. (Adapted from work done by Gaymes et al).

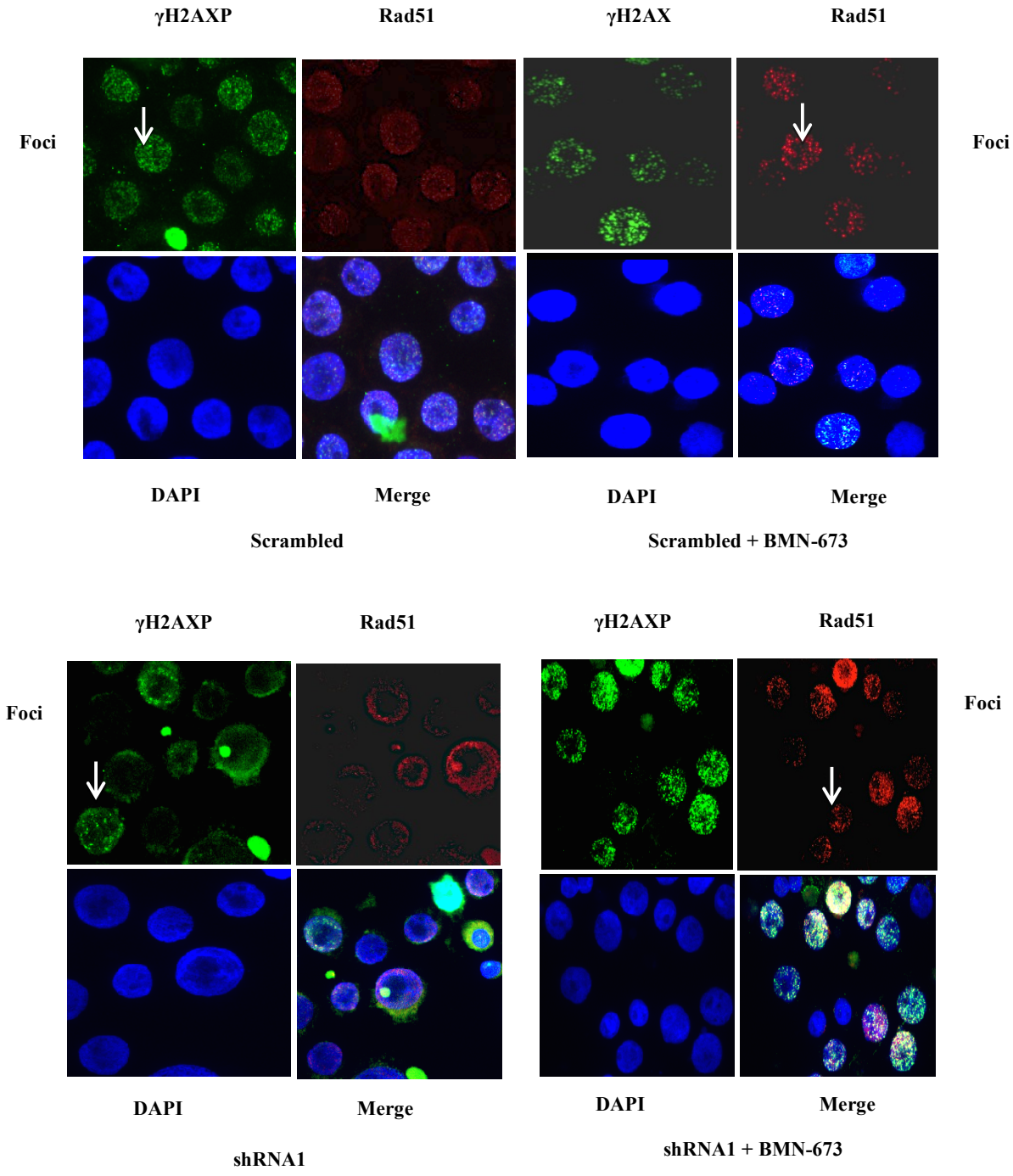


Figure 4.12: Effect of PARP inhibitors on *STAG2* knockdown U-937 cell line. Immunostaining of *STAG2* knockdown U-937 nuclei treated with 100nM BMN-673 for 24hr. Cells were probed for γH2AX-P foci (green stain) Rad51 foci (red stain) and nuclei stained with DAPI (blue) white arrows depicts the foci. The figure shows the frequency of cells displaying Rad51 foci (%) (red bars) γH2AX-P (%) (green bars) as detected by immunofluorescence following addition of BMN-673. 200 nuclei were counted per experiment n=2.

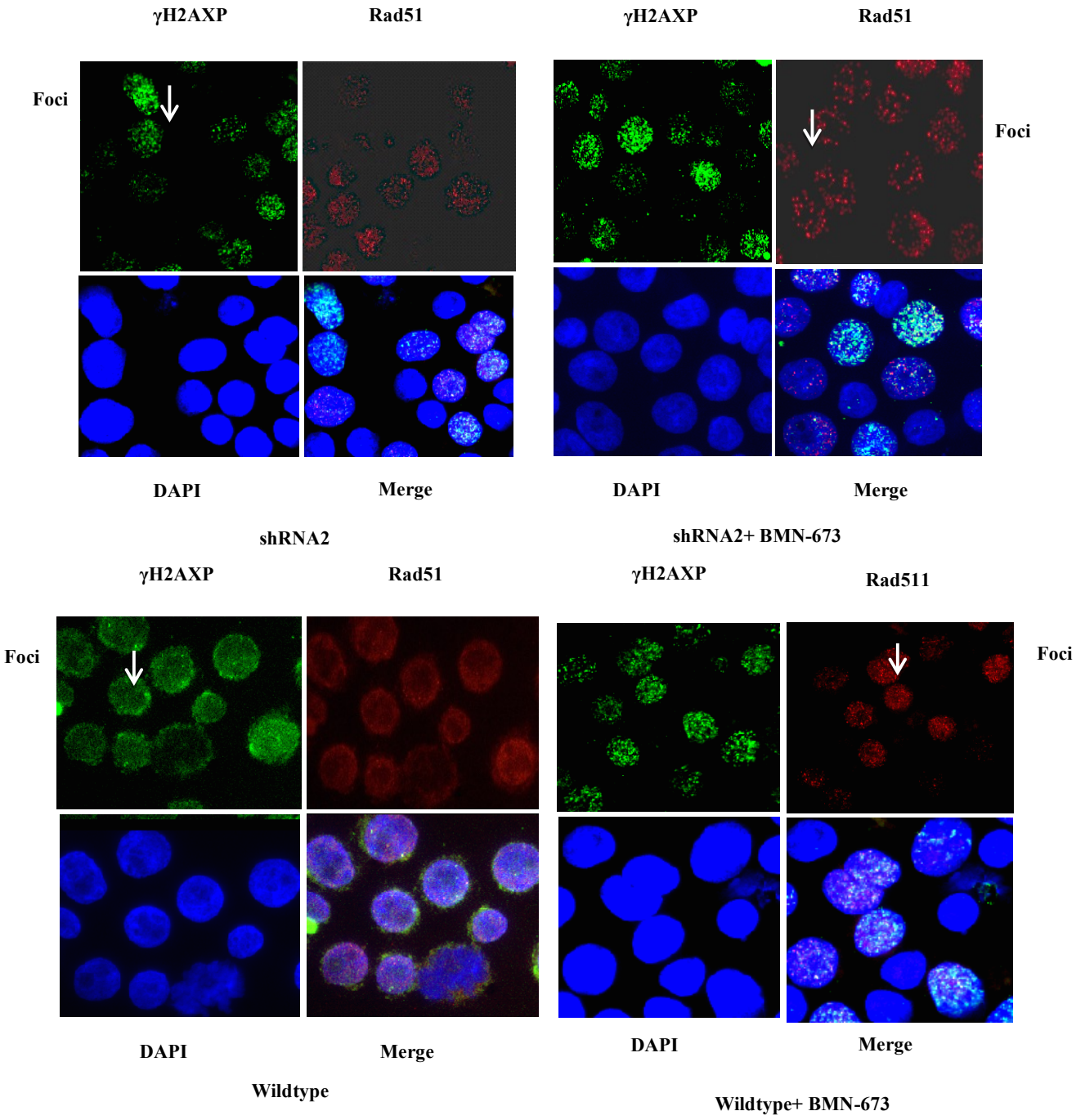


Figure 4.13: Effect of PARP inhibitors on *STAG2* knockdown U-937 cell line. Immunostaining of *STAG2* knockdown U-937 nuclei treated with 100nM BMN-673 for 24hr. Cells were probed for γH2AX-P foci (green stain) Rad51 foci (red stain) and nuclei stained with DAPI (blue) white arrows depict the foci. The figure shows the frequency of cells displaying Rad51 foci (%) (red bars) γH2AX-P (%) (green bars) as detected by immunofluorescence following addition of BMN-673. 200 nuclei were counted per experiment n=2.

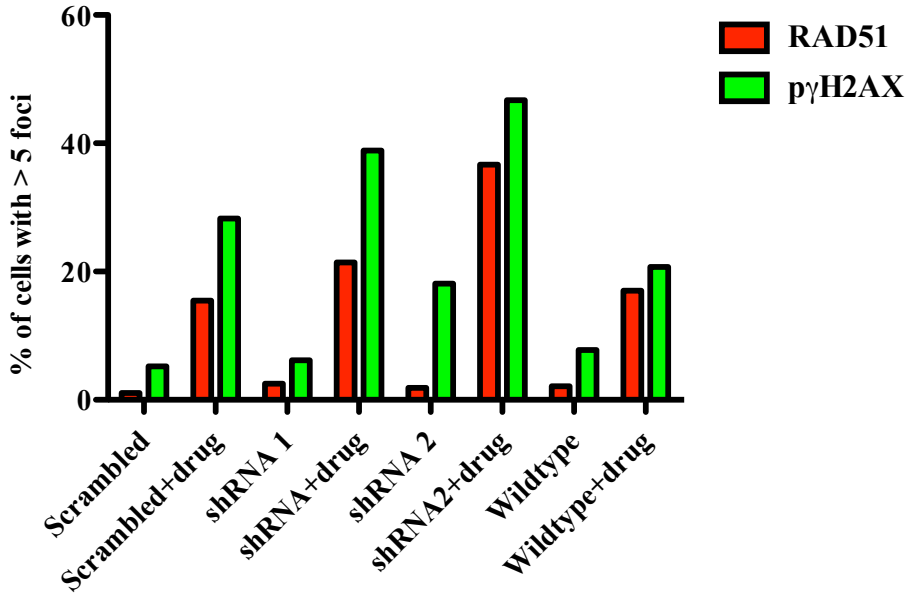


Figure 4.14: Percentage of cells displaying Rad51 foci (red bars) γH2AX-P (green bars) as detected by immunofluorescence following addition of BMN-673. 200 nuclei were counted per experiment. n=2.

Increase in phosphorylation of H2AX in the form of foci was observed on addition of BMN-673 to knockdown cells as compared to treated scrambled and wildtype cells however this increase was not accompained by loss of RAD21 foci.

4.5.5 Effect of PARPi on *STAG2* deficient cell viability

To evaluate the additive effect of PARP inhibition and *STAG2* deficiency on cell viability. *STAG2* shRNA treated U-937 were exposed to BMN-673 and cell counts were performed on days 2, 7 and 12 using the trypan blue exclusion method (Fig 4.15). Trypan blue is a vital stain that distinguishes between live and dead cells. It is readily taken up by dead cells whose membrane is not intact, in contrast to viable cells.

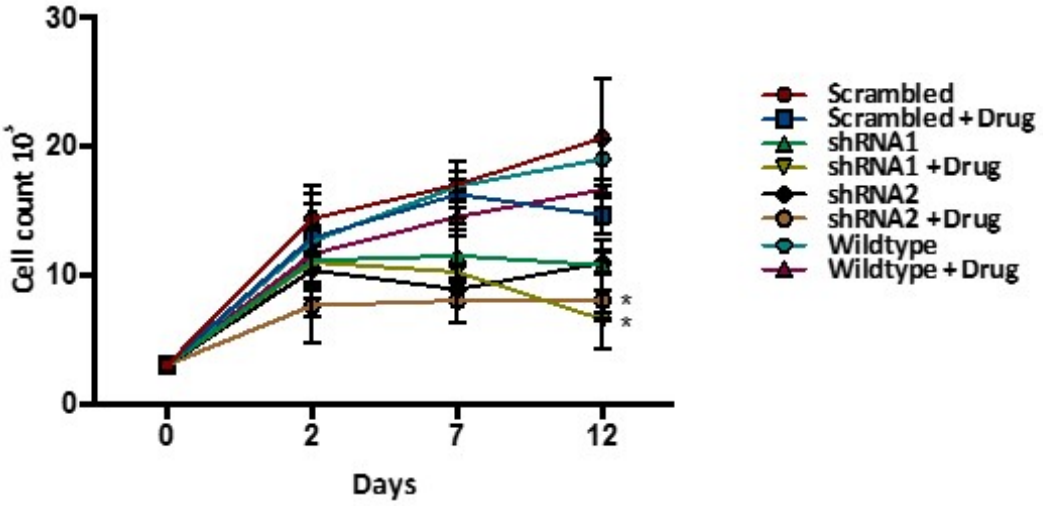


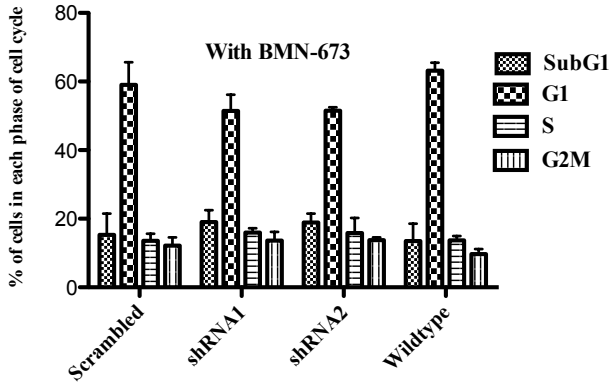
Figure 4.15: Cell survival of *STAG2* silenced U-937 treated with BMN-673. *STAG2* silenced. U-937 cell line was exposed to BMN-673 for 12days and cell count was determined using trypan blue exclusion analysis. *p<. 05 for both shRNA1 and shRNA2 *STAG2* deficient cells treated with BMN-673 as compared to control cells using trypan blue exclusion method however no cell death was observed.

Significant difference in cell count was observed between treated and untreated shRNA1 and shRNA2. *STAG2* deficient cells but low cell count was not associated with any cell death although increase in phosphorylation of H2AX was observed. This result can be supported by the fact that under the influence of drug, treated cells need more time to repair as compared to untreated cells hence resulting in slower cell growth.

4.5.6 Cell cycle

Cell cycle was performed on *STAG2* knockdown cells exposed to BMN-673 for 12 days. Cells were stained with FITC and PI (and cell cycle was performed using flow cytometry). The percentage of cells in each phase of the cycle was compared between *STAG2* silenced U-937 cells treated with the PARPi and untreated cells (Fig 4.16).

(a)



(b)

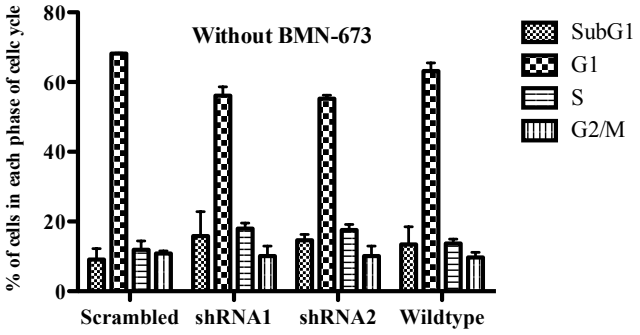


Figure 4.16: Effect of BMN-673 on cell cycle in shRNA silenced U-937 cell line. Knockdown cells were exposed to BMN-673 for 12 days (a) treated and (b) untreated cells were stained with fluorescence isothiocyanate (FITC) and propidium iodide and then analyzed using flow cytometry. Although a small increase in the percentage of cells in the subG_1 phase for both treated shRNA1 and shRNA2 could be observed the decrease was insignificant.

The cell cycle profile did not vary significantly in different phases of the cell cycle between BMN-673 treated and untreated shRNA silenced *STAG2* U-937. Moreover, no significant differences were observed in sub G1 (apoptotic phase) with or without BMN-673.

4.5.7 Annexin V assay

The Annexin V assay was performed at day 12 to support the data from the cell viability and cell cycle assays (Refer to section 2.6 for more details). Annexin V assay determines the degree of apoptosis via binding of annexin to phosphatidyal serine (PS) on the surface of apoptotic cells (Fig 4.17).

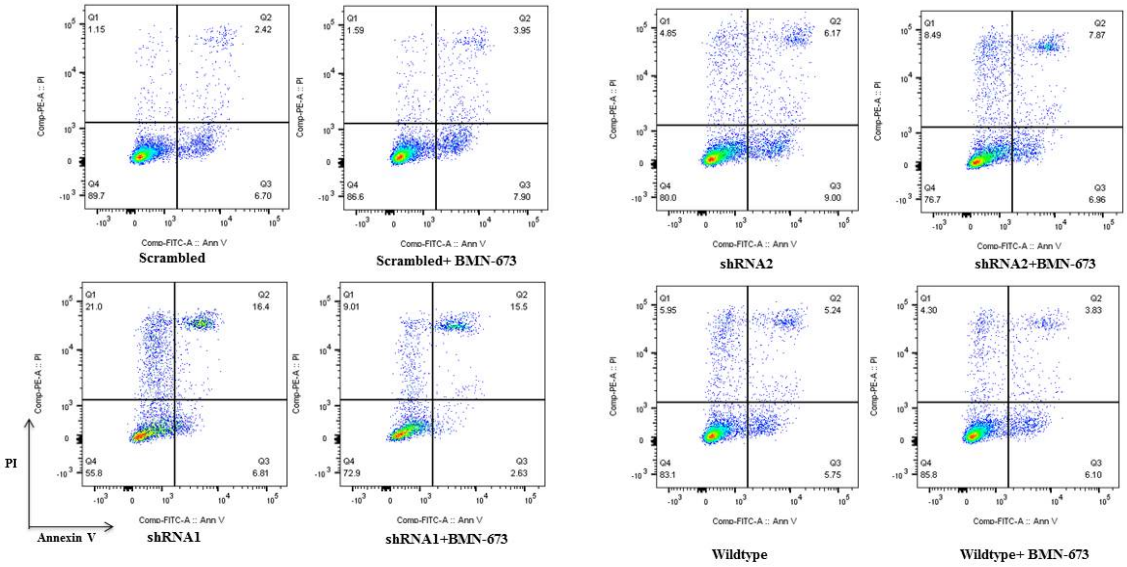


Figure 4.17: Annexin V assay on *STAG2* shRNA treated U-937 treated with 100nM of BMN-673. Annexin V and PI staining of shRNA transduced U-937 cells treated with BMN-673 at day 12. Q1 upper left quadrant represents cells + for PI. Q2 upper right quadrant represents cells double +ve for Annexin V and PI. Q3 bottom right quadrant represents cells +ve for Annexin V and Q4 bottom left quadrant represents cells double –ve for Annexin V and PI.

Although increase in apoptotic cells was observed in shRNA1 treated cells this increase was not significant as compared to untreated knockdown cells. No difference in apoptotic cells was found between treated *STAG2* silenced cells as compared to untreated controls.

Annexin V assay supported the findings from viability and cell cycle assay as no difference was found in the percentage of early apoptotic and late apoptotic cells between the treated vs untreated shRNA transduced U-937 cells.

4.6 The UMUC3 cell line

4.6.1 Evaluation of Expression of STAG2 in *STAG2* mutated cell line UMUC3 by Western Blotting

UMUC3 is an adherent hypertriploid urinary bladder cell line. It has a single base deletion that creates a nonsense mutation at position K983X resulting in no expression of STAG2 as shown in (Fig 4.18 a). Loss of the endogenous allele of *STAG2* formed the basis of selecting this cell line to determine the functional consequences of complete loss of STAG2. The effect of the complete loss of STAG2 expression on other members of the cohesin complex was also determined (Fig 4.18 b).

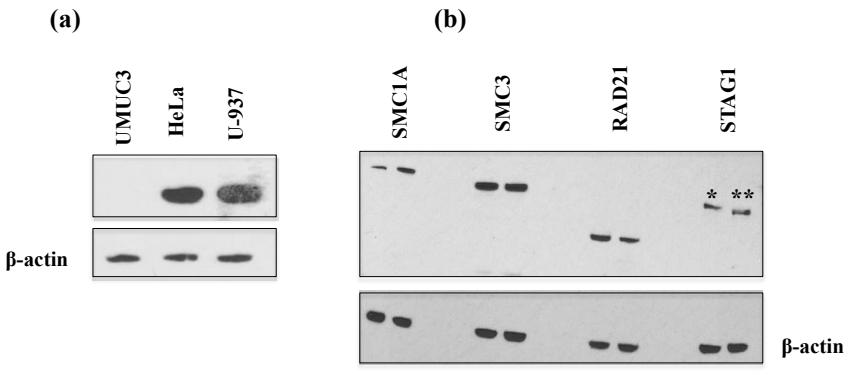


Figure 4.18: Expression levels of members of cohesin complex in UMUC3 cell line. (a) Evaluation of STAG2 expression in UMUC3 by western blotting, compared with Hela and U-937 cell lines used as controls . Loss of expression of STAG2 was confirmed in UMUC3 cell line. β -actin was used as a loading control (b) Expression of other members of the cohesin complex SMC1A, SMC3, RAD21 and STAG1 in UMUC3* and U-937 were also evaluated no difference in expression was observed between the cell lines.**

Western blot analysis confirmed complete loss of STAG2 expression in UMUC3 cell line however the loss in STAG2 expression was not accompanied by concomitant decrease in expression of the other members of the cohesin complex.

4.6.2 Homologous Recombination Assay

Homologous Recombination assay was performed on UMUC3 to evaluate whether complete loss of STAG2 influences HR repair. UMUC3 was exposed to 100nM BMN-673 for 24hr and then immunostaining was carried out. UMUC3 cells were probed for RAD51 (marker for HR) and γH2AX-P (marker for DNA damage) (Fig 4.19and 4.20).

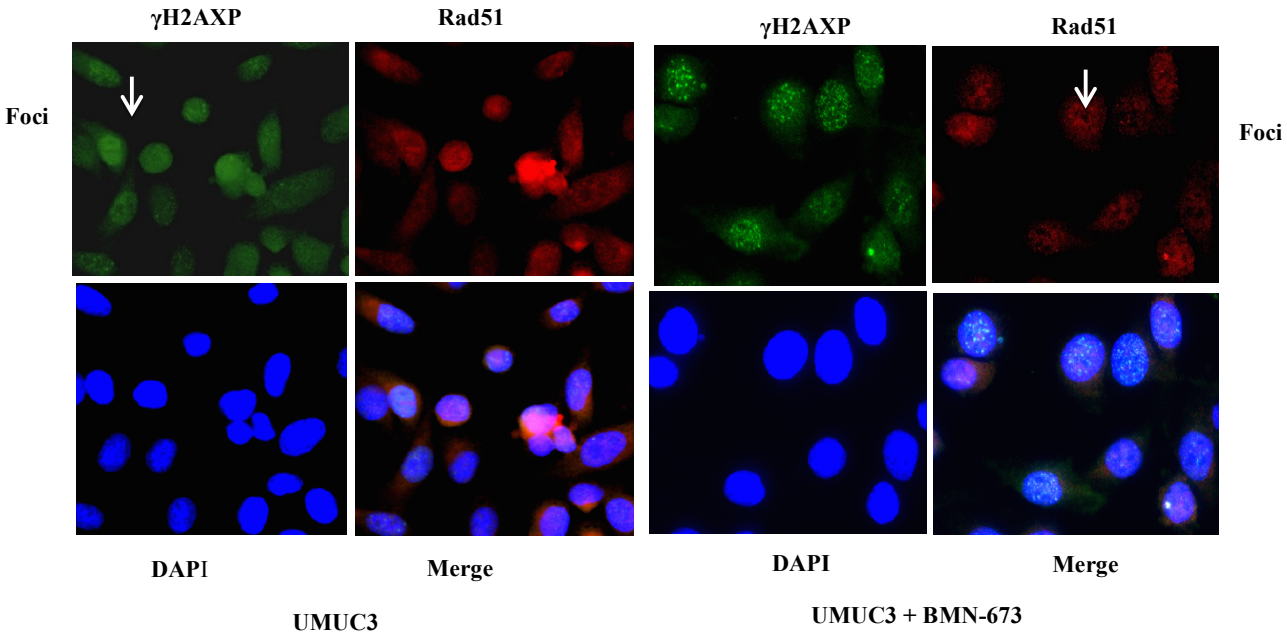


Figure 4.19: Effect of PARPi on cell line (UMUC3) with no expression of *STAG2*. Immunostaining of UMUC3 nuclei treated with 100nM BMN-673 for 24hr. Cells were probed for Rad51 foci (red stain) γH2AX-P foci (green stain) and nuclei stained with DAPI (blue). White arrow represents the foci.

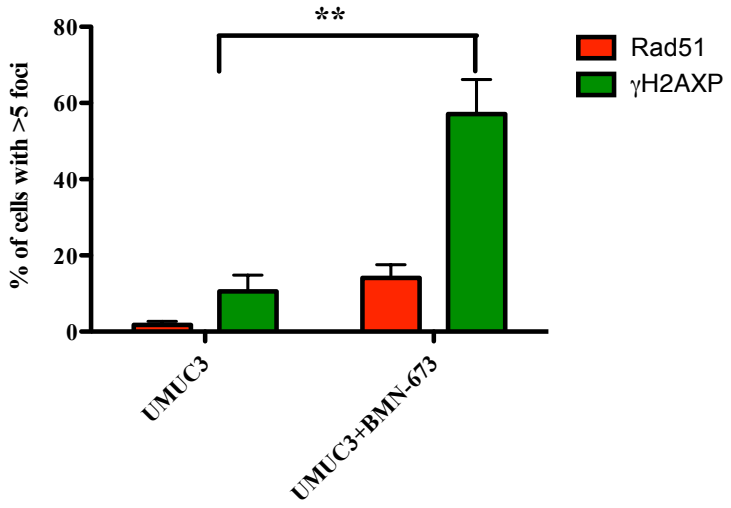


Figure 4.20: Percentage of cells displaying Rad51 foci (red bars) γH2AX-P (green bars) as detected by immunofluorescence following addition of BMN-673. 200 nuclei were counted per experiment. n=3. Significant difference was found in the phosphorylation of γH2AX. **p<. 01 between the treated and untreated cells.

Treatment of UMUC3 with 100nM BMN-673 resulted in a significant increase in the number of γH2AXP foci but this was not accompanied by any cell death further lack of STAG2 expression had no effect on recruitment of RAD51 foci although the number of foci observed were less as compared to that observed in treated *STAG2* knockdown U-937 cell line.

To study the effect of BMN-673 on UMUC3 viability cells were treated with BMN-673 for 12 days and cell counts were performed at fixed time points using trypan blue exclusion MCF-7 cell line an adherent breast cancer cell line with expression of STAG2 was used as a control (Refer to section 2.1.2 for more details) (Fig 4.21).

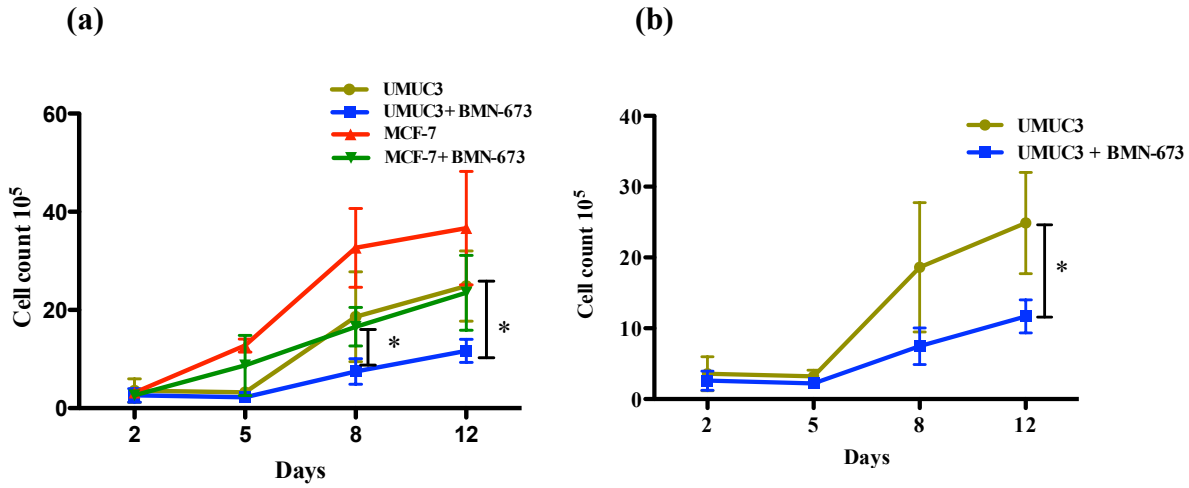


Figure 4.21: Cytotoxicity curves of treated UMUC3 cell line. (a) Cell counts were performed on day 2, 5, 8 and 12 using the trypan blue exclusion assay. MCF-7 was used as a control. Significant difference was found between treated and untreated UMUC3 at day 12 and between treated UMUC3 and MCF-7 at day 8 p*<. 05. Data presented as mean ±SD (n=3). (b) A separate graphical plot for day 12 between treated and untreated UMUC3 cell.

Significant difference in cell count was found between treated and untreated UMUC3 cell line on day 8 and between treated UMUC3 and MCF-7 cell line on day 12 but this decrease in cell number was not associated with cell death**.**

4.6.3 Cell cycle

To study the impact of BMN673 treatment on UMUC3 cell cycle kinetics, UMUC3 cells were exposed to 100nM BMN-673 for 12 days. Cells were collected and stained with FITC and PI and analysed using flowjo (Fig 4.22).

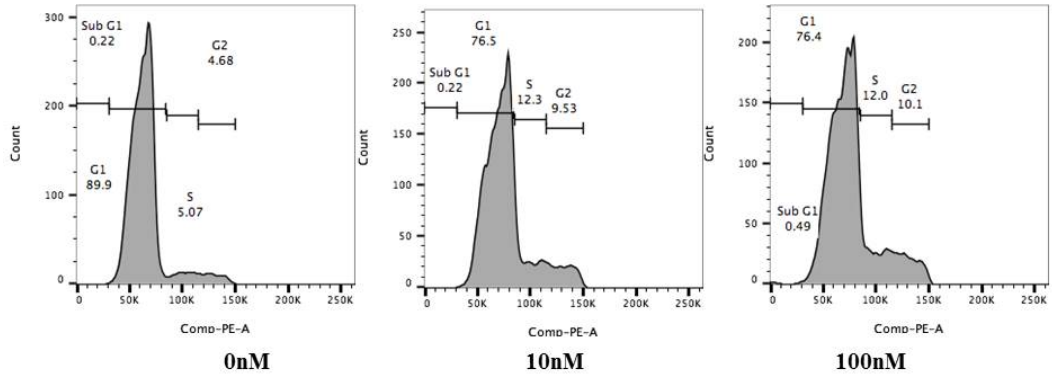


Figure 4.22: Cell cycle profile of UMUC3 cell line treated with 100nM of BMN-673 for 12 days. Cells were exposed to increasing concentrations of BMN-673 for 12 days and cell cycle kinetics was determined using flow cytometry. No significant difference was found in the percentage of cells in different phases of cell cycle.

No effect was observed in any phases of cell cycle after treating UMUC3 cell line with increasing concentrations of BMN-673, was in agreement with the data obtained from *STAG2* shRNA silencing of cell cycle kinectics**.**

4.6.4 Annexin V

Annexin V analysis was performed on UMUC3 exposed to 10 nM and 100 nM of BMN-673 for 12 days (Fig 4.23).

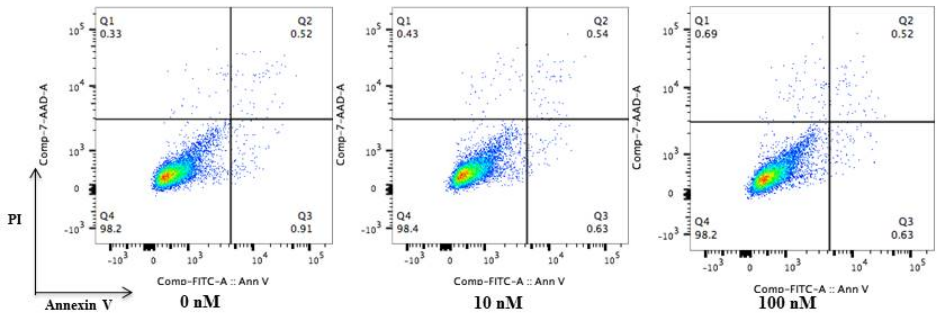


Figure 4.23: Annexin V analysis on UMUC3 cell line treated with 10 and 100nM BMN-673 continuously for 12 days. At day 12 Annexin V assay was performed. The upper left quadrant Q1 represents –ve cells for 7AAD, the upper right quadrant Q2 represents double + for FITC and 7AAD V, the lower left quadrant represents double –ve for both Annexin V and 7AAD and lower right quadrant represents +ve for Annexin V.

No difference in early apoptotic and late apoptotic cells was observed in UMUC3 cells treated with increasing concentrations of BMN-673. Based on the results obtained from HR assay, viability, cell cycle and Annexin V one can infer that loss of expression of *STAG2* has no effect on homologous recombination and further STAG2 deficient cells are not sensitive to PARPi.

4.6.5 Synergistic effect of Cisplatin and Aurora Kinase B with BMN-673

No enhancement in cytotoxicity was observed with combined treatment of BMN-673 with Cisplatin (a DNA damaging agent) and Aurora kinase B after treating the cells for 19hr. Aurora kinase B inhibitor and cisplatin were chosen to determine if synergistic inhibition using Aurora kinase B (involved in phosphorylation of sororin) or cisplatin with PARPi could enhance the cytotoxicity in *STAG2* deficient cells. Although treating knockdown cells individually with 1μg/ml cisplatin and 20nM Aurora kinase B showed cytotoxicity at 19hr but no enhancement in cytotoxicity was observed at 48hr timepoint (Fig 4.24).

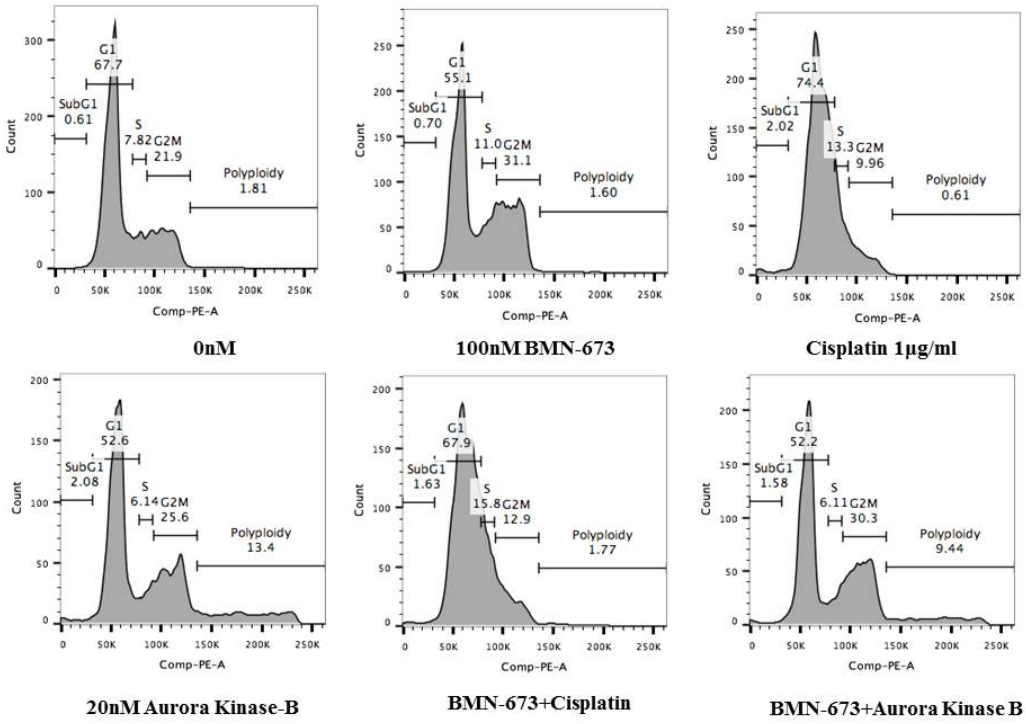


Figure 4.24: Synergistic inhibitory effect of Aurora kinase B and Cisplatin with BMN-673 on *STAG2* deficient UMUC3 cell line after 19hrs. UMUC3 cell line was treated with BMN-673, Aurora Kinase B and Cisplatin for 19hr. Treated cells were stained with PI and FITC and effect on cell cycle kinetics was determined by running the samples on BD FACS Canto and result analyzed using flow jo.

4.7 Discussion

Apart from its established role in maintaining the polarity of sister chromatids during mitosis, cohesin is also involved in a large number of biological processes such as DNA damage repair and gene regulation (Panigrahi and Pati, 2012). A major role of cohesin as a tumor suppressor gene has been revealed by genomic and clinical analysis, but its link with leukaemogenesis is still undermined. The majority of the cohesin mutations in myeloid malignancies such as MDS and AML are heterozygous mutations and are mutually exclusive.

Here, *in-vitro* analysis was performed to delineate the role of cohesin complex in leukaemic progression by silencing *STAG2*; a subunit of the cohesin complex in myeloid and urinary bladder cell lines.

STAG2 is a gene located on X chromosome and majority of the mutations are nonsense mutations resulting in a truncated protein. U-937 is a malignant cell line that was derived first from a pleural effusion of 37 year old Caucasian male with diffuse histiocytic lymphoma. It is one of the only few human cell lines still expressing many of the monocytic like characteristics exhibited by cells of histiocytic origin.

In order to determine the effects of *STAG2* downregulation, we used siRNA from Dharmacon and transfected cell samples were collected at various timepoints. Using qRT-PCR, maximum silencing was achieved at 24hr timepoint, with 67% reduction. Expression levels of *STAG1,* a paralog of *STAG2* involved in telomere replication was also evaluated as previous reports have indicated compensatory elevation of *STAG1* upon knockdown of *STAG2.* However, no increase in the expression level of *STAG1* was observed upon knockdown of *STAG2* as confirmed by qRT-PCR. In order to increase the knockdown efficiency and authenticity of

knockdown, another siRNA from Santa Cruz was used. However, there was no increase in knockdown efficiency or increase in the level of *STAG*1 expression. Cell cycle samples collected upto 96hr also did not show any significant difference between the *STAG2* knockdown and control samples.

A connection between cohesin mutations and cell cycle irregularities has been investigated by various groups including work done by *Mazumdar et al* where the role of cohesin mutations was studied in the differentiation and self renewal of human haematopoietic stem and progenitor cells (HSPCs). The group engineered human erythroid and AML cell lines to express wildtype or mutant cohesin (including missense and nonsense mutations) in all four cohesin components under the control of doxycycline -inducible promoter and found no significant changes in proliferation or cell death upon knockdown of cohesin sub-units (Mazumdar et al., 2015). The lack of discernible phenotype is supported by other studies demonstrating that during metaphase only a very small fraction of cohesin is required to hold the sister chromatids together and as knockdown of *STAG2* was not 100% in our analysis, the residual STAG2 was sufficient to hold sister chromatids together and as such no effect on cell cycle was observed. Also study carried out by *Kim et al* has shown that STAG2 is not involved in ring formation (Kim et al., 2016).

The transient nature of siRNA knockdown could be one of the reasons behind the results obtained from cell cycle and potentially could be resolved by generating a stable knockdown cell line. Unsuccessful attempts to generate a stable knockdown cell line using the retroviral expression vectors prompted the application of lentiviral mediated knockdown of *STAG2*. Two pre-designed shRNAs from Addgene were used which gave a knockdown efficiency of 72 and 76% that was

confirmed at the mRNA level by qRT-PCR and protein level by western blotting. Expression levels of other core cohesin subunits was also determined by western blotting. No changes in the expression level of other members of the cohesin complex were observed upon shRNA knockdown of *STAG2*. Work done by *Bailey et al* in two glioblastoma cell lines, with an inherent defect in *STAG2* also demonstrated no increase in the expression level of other cohesin components in *STAG2* mutated cell lines, further supporting my data (Bailey et al., 2013). Conventional chemotherapeutics agents cannot differentiate between the normal and cancerous cells resulting in the death of both. Up until now, targeted therapy that is selective for the malignant clone has been an unachievable target, however the advent of targeted therapeutics in clinical use illustrates the enormous potential of this approach. The concept of synthetic lethality originates from work in Drosophila where a combination of mutations in two genes is deleterious for the cell in contrast to mutations in one gene which restores viability. Changes in the genetic makeup of cancer cells has been delineated by a large number of studies that make them a target for synthetic lethality approaches. Synthetic lethality induced by PARPi has given a renewed enthusiasm since its application in breast cancer susceptibility *BRCA1* and *BRCA*2 mutated cancers to develop anticancer agents that can target cancer cells while sparing normal tissues (Bryant et al., 2005, Farmer et al., 2005). Although different underlying mechanisms have been proposed they are mostly attributed to critical functions of PARP in variety of DNA processes including as a sensor of single strand breaks (SSBs) in base excision repair (BER), acting as a mediator to restart stalled replication forks through HR mediated double strand break (DSB) repair and preventing binding of Ku to DNA ends in non homologous end joining (NHEJ) (Bryant et al., 2009, Haince et al.,

2008, Paddock et al., 2011). Double strand DNA damage within the cell is repaired by two different pathways, error free homologous recombination (HR) pathway and the error prone NHEJ pathway. HR pathway predominates in mid-S and mid-G_2 phase where a sister chromatid is available. Helicases and endonucleases carry out extensive resection generating long stretches of ssDNA and interaction of various proteins i.e. ATM, MRE11, CtiP, BRCA1 and BRCA2 play an important role in repair. One of the most important proteins involved in DSB DNA repair is the recombinase RAD51, which forms a nucleofilament on ssDNA that drives strand exchange with a homologous template strand. As cohesin is involved in DNA damage repair, we hypothesized that cohesin i.e. STAG2 may be involved in HR repair and hence treatment with PARPi in the background of *STAG2* silencing could result in increased cytotoxicity and cell death and the establishment of synthetic lethality between STAG2 and PARP. This hypothesis was supported by work done by *Gaymes et al* who showed that leukaemic cell lines were sensitive to PARPi, PJ-34, KU-58948 (now known as Olaparib), and EB47. The cell cycle profile of P-39 showed an increase in number of cells in S phase and G2/M phase indicative of stalled replication and increased DNA damage. Immunocytochemical analysis showed that the cell lines, P-39, Mutz-3 and primary cells from two AML patients showed increased ϒH2AXP foci and decreased RAD51 foci as compare to resistant cell lines and primary AML patients suggesting that the sensitivity to PARPi was due to a defect in Homologous Recombination in these cells (Gaymes et al., 2009). BMN-673 (Talazoparib), a highly potent PARPi was used in my analysis. With an IC_{50} value of 0.57nM it significantly inhibits H_2O_2-elicited PAR synthesis in vitro (Wang et al., 2016).The high potency of BMN673 is perhaps due to its PARP trapping capacity at SSB sites. *STAG2* knockdown cells treated with the inhibitor

showed an increase in ϒH2AXP foci as compared to control but recruitment of RAD51 to areas of DNA damage could also be observed. This result suggests that STAG2 is not involved in homologous recombination pathway i.e. recruitment of RAD51. Studies have also shown that other factors can be responsible for suppression of homologous genes and can indirectly confer sensitivity to PARP inhibitors. A study carried out by *Esposito et al* demonstrated that sensitivity of repressive transcriptional factors like AML1-ETO (encoded by the fusion oncogene RUNX1-RUNX1T1) and PML-RARα fusion oncoproteins (encoded by PML-RARA) to PARPi is due to suppression of homologous repair genes and inhibition of recruitment of RAD51 however overexpression of Hoxa9 results in loss of sensitivity incontrast tO mixed lineage leukaemia (MLL) which are resistant to PARPi is due to Hoxa9 and inhibition of Hoxa9 rescues this phenomena and confers MLL sensitivity to PARPi (Esposito et al., 2015).

The effect of *STAG2* knockdown on viability was studied by exposing cells to 100nM of BMN-673 for 12 days. The knockdown cells treated with 100nM BMN-673 showed significant decrease in cell count compared to control cells by trypan blue exclusion assay but no dead cells were observed during counting. One of the reasons for low cell count can be attributed to the fact that treated cells unlike untreated require more time to repair the damaged caused by PARP inhibitors hence resulting in slow proliferation and accounting for low cell count

Cell cycle kinetics was performed on *STAG2* silenced U-937 that were exposed continuously to PARPi for 12 days. No significant difference in the percentage of cells in each phase of the cell cycle was observed. This result is in contrast to results obtained by *Bailey et al* where treatment of glioblastoma cell lines over a wide range of olaparib concentration showed a significant decrease in cell count as

compared to their knock in counterparts. Treatment of *STAG2* mutated cells with olaparib showed accumulation of cells in G_2 phase, increase in micronuclei, chromatin bridges (Bailey et al., 2013). This discrepancy might be due to the fact that the shRNA knockdown of U-937 did not result in complete loss of STAG2 protein in my study in contrast to work done by Bailey et al where both the cell lines had complete loss of STAG2 expression. In my study, the residual *STAG2* could be sufficient enough to hold the sister chromatids together as a result of which no synthetic lethality was observed between STAG2 and PARPi. In a recent study, comparison of karyotypic analysis of c-Kit+ progenitor cells in $Rad21^{(shRNA/+)}$ $Smc1a^{(shRNA/+)}$ and $Stag2^{(shRNA/+)}$ animals demonstrated no gross chromosomal instability. The report supported my observation by the fact that despite efficient knockdown of Rad21, Smc1a or Stag2, the remaining protein will be sufficient to form a functional cohesin complex (Mullenders et al., 2015).

UMUC3 is an adherent urinary bladder cell line. It has a single base deletion that creates a nonsense mutation at position K983X. This cell line was chosen as it had no expression of STAG2. Loss of STAG2 had no effect on the expression level of other subunits of the cohesin complex nor upregulation of STAG1. Treatment with 100nM of PARPi BMN-673 resulted in increase in phosphorylation of histone variant ϒH2A.X at Ser 139. Although recruitment of RAD51 was observed foci formation was decreased as compared to *STAG2* knockdown U-937 cell line. UMUC3 cell line exposed to 100nM PARPi for 12 days showed significant lower cell count as compared to untreated UMUC3 as well as treated MCF-7 cell line used as a control however decrease in cell number was not accompanied by cell death. The observed effect could be due to the fact that treated cells unlike untreated cells take more time to repair the damaged caused by the inhibitor than and hence

account for slower proliferation while different genetic background between UMUC3 cell line (adherent urinary bladder cell line) and MCF-7 cell line (adherent breast cancer cell line) could be the reason for the effect observed between them. No effect on cell cycle was observed on treatment of UMUC3 cell line with two different concentrations of PARPi for 12 days which was supported by Annexin V results. No sensitivity to PARPi was observed in UMUC3 cell line despite having no expression of STAG2. The results obtained from UMUC3 adherent cell line as well as those obtained from *STAG2* silenced U-937 myeloid cell line convey that *STAG2* in such a setting is not involved in HR pathway and further loss of *STAG2* does not contribute towards sensitivity to PARPi.

To enhance the cytotoxicity of PARPi BMN-673 UMUC3 cell line was treated with DNA damaging agents cisplatin and Aurora Kinase B inhibitor. Cisplatin is a chemotherapy drug which is the first member of platinum containing anti-cancer agents. Cisplatin binds to DNA causing the DNA strands to cross link which ultimately results in cell death. Regulation of cell cycle transit from G2 through cytokinesis is regulated by Aurora Kinases. The scattered mitotic spindles formed due to mutant Aurora kinase have become a target for anti cancer drug development. No enhancement in cytoxicity was observed on combined treatment of BMN-673 with cisplatin and aurora kinase B inhibitor.

Cohesin complex has many roles in different pathways and DNA damage repair is among one of them. It can be hypothesized that mutations in cohesin complex can indirectly contribute to synthetic lethality by disregulating the function of a large number of genes involved in various other pathways like transcriptional regulation against which inhibitors can be designed. This can be supported by a study, which opened up the intriguing possibility that *IDH1/2* mutations contribute to

leukemogenesis not only through epigenetic changes but also through dysregulation of mithochondrial function where mutations in *IDH1*/*IDH2* result in sensitivity to anti-apoptotic genes like *BCL*. The mechanism behind this sensitivity was the accumulation of 2-hydroxyglutarate due to *IDH1/IDH2* mutations that inhibits enzymatic activity of complex IV of the mitochondria and prevents mitochondrial proapoptotic BCL-2 family members BAX and BAF to cause apoptosis by binding to BCL2. However inhibition of BCL2 by ABT-199 inhibitor prevents this interaction and lead to apoptosis (Chan et al., 2015).

Recent studies carried out by four different groups have strongly pointed towards a dominant and highly specific role of cohesin in regulating cell fate decisions in HSCs. The studies provide a solid platform to understand the role of cohesin mutations in myeloid malignancies.

Mazumdar et al showed that introduction of cohesin mutations in an erythoid (TF-1) and AML cell line (THP-1) and primary cord blood HSPC resulted in differentiation block. A similar feature was seen with knockdown of *RAD21* both *in vivo* and *in vitro*. Mutant cohesin lead to increased chromatin accessibility and a higher predicted binding at transcription factor binding sites to GATA2, ERG and RUNX1 (critical regulators of HSPCs). Knockdown of these transcription factors reverted the differentiation block induced by cohesin mutants. The cohesin mutants were also shown to impart its differentiation block only on the immature HSCs and MPPs (Multi potent progenitors) (Mazumdar et al., 2015).

Mullenders et al generated a series of shRNA silencing models in which endogenous cohesin could be inducibly reduced. The authors showed that silencing of cohesin complex members did not have a deleterious effect on cell viability. However, an increase in serial cell plating was observed in mouse bone marrow

haematopoietic progenitor cells. *In vivo*, cohesin knockdown had an effect on myelopoiesis and homeostasis. Furthermore, knockdown of Stag2 showed an increase in chromatin accessibility around genes that were found to be upregulated in *STAG2* knockdown cells such as Fc- ϒ receptor (Fcgr3, and Fcgr4) and decrease in chromatin accessibility around genes that were found to be downregulated such as CD74. Increase accessibility in GATA1 locus in Stag2 knockdown LSK cells was found. Finally, an aged mouse showed varying degrees of splenomegaly, myeloid hyperplasia and promoted a clinical picture consistent with myeloproliferative neoplasm (MPN) (Mullenders et al., 2015).

A conditional expression model of *SMC3 in vivo* was created to delineate the role of *SMC3* in haematopoietic function. The study showed a dose dependent requirement of *SMC3* in haematopoiesis. Biallelic loss of *SMC3* lead to 100% lethality, whilst heterozygous loss of *SMC3* tilted the balance towards self-renewal rather than lineage commitment. The authors concluded that cohesin mutations were associated with gene regulation and transcriptional output of hematopoietic cells rather than chromosomal instability. *SMC3* haploinsufficient cells also showed an increased capacity of serial replating and competitive advantage in bone marrow transplants in agreement with an enhancement of self renewal. The transcriptome of *SMC3* haploinsufficient mice showed a downregulation of certain genes such as *Pou2af1* and *Gfi1b* that are involved in lineage committed cells. *Smc3* monoallelic deletion cooperated with *FLT3-ITD* mutation to induce AML *in vivo,* thus indicating the potentiation of Stat5 signalling and increased expression of Stat5 targets. There was also a preponderance of multifocal nucleoli seen in $Smc3^{\Delta/+}/Flt3^{ITD}$ which resembled the nucleolar alteration seen in Roberts syndrome, a cohesinopathy caused by mutation in *ESCO2* that inactivates an acetyl transferase required to form

the cohesive structures. The authors also speculated that leukemogenesis arising due to cohesin complex mutations in in addition to disrupted transcription regulation may also arise due to defective protein synthesis as Robert syndrome nuclei also show disrupted ribosomal RNA expression and ribosome biogenesis (Viny *et al* 2015). *Galeev et al* made use of a genome wide RNAi screen and found members of the cohesin complex (*SMC3*, *RAD21*, *STAG2* and *STAG1*) among the top 20 genes whose knockdown maintained the HSC phenotype in culture. *STAG2* knockdown *in vivo* showed decreased differentiation and myeloid skewing as determined by an increase in CD33/CD15-positive cells. Knockdown of *SMC3* despite showing an increase in undifferentiated cells *in vitro* failed to demonstrate a skewing towards the myeloid lineage despite an increase in engraftment *in vivo*. Although several of the observed effects in cohesin knockdown models showed signs of myeloid neoplasms, any signs of malignant diseases in either primary or secondary transplanted mice were absent indicating that cohesin deficiency alone cannot trigger leukaemia. A number of HSC related genes that have a role in HSC regulation were identified in the gene expression profiling for STAG2 or SMC3 and included *ERG*, *EGR1*, *KLF5* and *SOCS2*. (Galeev et al., 2016).

In conclusion work carried out here demonstrate that *STAG2* is not involved in HR which has been demonstrated in the background setting of two cell line with minimal and complete loss of expression of STAG2. Further loss of STAG2 did not sensitize cells to PARP inhibitor BMN-673 and also to cisplatin and aurora kinase B inhibitor. Recent studies show that mutations in cohesin complex which includes *STAG2* as well contribute to leukaemogenesis through dysregulation of chromatin architecture thereby strengthening its role as a gene regulator or else like *IDH1/2* mutations *STAG2* mutations may also confer senstivity to inhibitors by affecting

genes involved in other pathways. Such genes need to be identified and inhibitors need to be designed against them.

Chapter Five

GENERAL DISCUSION

The landscape of MDS is characterized by aberrations in genes involved in various signalling pathway that contribute towards pathogenesis of MDS. Mutations in the cohesin complex particularly *STAG2* a major component of the cohesin complex has been found in 5-15% of MDS cases. Mutations have also been identified in *SMC1A*, *SMC3*, *RAD21* besides various other regulatory subunits.

Although cohesin mutations tend to occur in myeloid malignancies the exact mechanism through which they contribute towards MDS is not clear. Chromosomal instability (CIN) is one of the hallmarks of human neoplasms. Despite its widespread prevalence, knowledge of the mechanisms and contributions of CIN in cancer has been elusive. Earlier studies associated cohesin mutations with chromosomal instability and aneuploidy however since then a significant number of publications disregard any association between cohesin mutations and aneuploidy and in fact most of the aberrations in cohesin mutant tumors are found to be euploid. p53 is considered as a marker for chromosomal instability and in concordance with several published reports in this study none of the patients with cohesin mutations has a p53 mutation and five out of nine patients has normal cytogenetics thus disregarding any association with chromosomal instability.

Most of the mutations in cohesin complex are nonsense and missense mutations. Due to the essential role of cohesin complex in sister chromatid cohesion (SCC) complete loss of cohesin has been proven to be incompatible with proliferation and cell survival and hence all aberrations are found to be monoallelic and never caused by a complete loss or multiple heterozygous mutations. Similar alteration in genotypic architecture caused by cohesin mutations is reported in my study. Majority of the mutations in my cohort are nonsense mutations and none of the cohesin mutations are found to co-exist together. The genes coding for *SMC1A* and *STAG2* are located on X chromosome and

are expected to have a stronger impact as no wild type is present however it has been reported that very little cohesin is required to hold the sister chromatids together due to which no effect on proliferation has been reported which is also evident from my study as transient as well as stable knockdown of *STAG2* has no effect on cell cycle kinetics as knockdown is not completely 100%.

Existence of genetically divergent subclones is a common feature of haematological malignancies. Less fit subclones tend to diminish with disease progression and only fit subclone dominate or various subclones can persist along with the dominant clones and aid in disease progression. In this study mutations in cohesin complex are found to coexist with mutations in epigenetic modifier *ASXL1* and splicesome components *SRSF2*. Although cohesin mutations have been reported to occur as subclones in the setting of co-occuring genetic and molecular abnormalities like *ASXL1* and key transcriptional regulators like (RUNX1, Ras family genes) this phenomenon could not be determined here as samples from patients were available only at the time of diagnosis and no follow up samples could be obtained however studies carried out have shown that in an individual setting mutations in *ASXL1* and cohesin genes, alter expression of Hoxa cluster and as cancer is a disease where mutations in genes involved in various pathways co-operate towards leukaemogenesis it can be speculated that both *ASXL1* and cohesin mutations in their altered state mutually co-operate to alter the functions of Hoxa gene cluster and hence contribute towards cancer progression (Abdel-Wahab et al., 2012, Fisher et al., 2016). Mutations in members of the cohesin complex have been found to co-exist with mutation in spliceosome component *SF3B1*. A study carried out demonstrated that mutations in splicesome component *SF3B1* resulted in missplicing of cohesin components and result in nonsense mediated decay linking this with my study it can be postulated that

mutation in *SRSF2* can result in missplicing of cohesin subunits and result in non-sense mediated decay however the loss of expression of cohesin subunits could not be confirmed at mRNA or protein level due to lack of viable cells from these patients (Sundaramoorthy et al., 2014)

Apart from holding the sister chromatids together cohesin is involved in various other functions due to its ability to form loops. Cohesin complex along with CTCF is responsible for maintenance of genomic architecture by forming tandem associated domains (TAD). Five of the mutations in *STAG2* are found in region were *STAG2* interacts with CTCF and loss of this interaction may have an effect on genomic architecture.

Mutations in cohesin complex are linked with various developmental disorders collectively called as cohesinopathies. The underlying etiology of these disorders are different while Cornelia de Lange Syndrome is caused by alteration in gene expression Roberts syndrome is a consequence of defective biogenesis. In this study one of the missense mutation found in *SMC1A* in one of the patient has been associated with Cornelia de Lange Syndrome. The mutation has been shown to impair intra or intermolecular approximation of the SMC head domains or disrupt binding of accessory proteins to the cohesin ring which can ultimately lead to developmental disorder (Deardorff et al., 2007).

The concept of synthetic lethality has been exploited to develop therapeutic agents that are specific for the maladaptive genetic changes in the cancer cell. The development of PARP inhibitors has accelerated rapidly after patients with heavily treated *BRCA1* and *BRCA2* mutations showed efficacy and feasibility to PARP inhibitors. In order to determine if *STAG2* has a role in HR a stable knockdown of *STAG2* was established in U-937 myeloid cell line using two shRNA and treated with

100nM of PARPi BMN-673 for 24hr. No effect on recruitment of RAD21 (a recombinase and a key marker of HR) has been observed leading to conclusion that *STAG2* is not involved in HR or the residual *STAG2* (as knockdown is not 100%) is be sufficient for effective HR. No effect on expression of other members of the cohesin complex, cell cycle, cell viability, and apoptosis is observed. The phenotypic effect of *STAG2* is also confirmed in a urinary bladder cancer line UMUC3 that despite having no expression of STAG2 still shows similar functional characteristics. A recent study has reported that STAG2 is not involved in the formation of ring and absence of STAG2 does not affect the ring formation ability this could be one of the reasons for the observed phenotype. Synthetic lethality has been observed between *PARP* and *STAG2* in glioblastoma cell line and also between cohesin and replication fork mediators. Lack of synthetic lethality between PARP and STAG2 in my study can be a cell specific effect as these studies were carried out in glioblastoma and cervical cancer cell line. Further no enhancement in cytotoxicity was observed in UMUC3 cell line with combinatorial effect of BMN-673 with cisplatin and aurora kinase B. It can also be possible that cohesin mutations in a myeloid setting contribute towards leukaemogenesis by altering the expression of genes involved in various pathways that need to be identified and inhibitors need to be designed against them.

Four studies undertaken to delineate the functional role of cohesin in murine and human haemopoiesis have all shown a similar pattern of preservation of immature phenotype of HSCs, impaired or delayed differentiation and skewing towards a myeloid lineage as well as deregulated gene expression pattern commonly associated with a preserved HSC phenotype and reduced activation of lineage specific programme. All these studies pointed out towards alteration in the cis regulatory chromatin architecture affecting gene transcription as the primary driver of cohesin

deficient architecture. Development of these mouse models will aid in understanding the molecular etiology of this complex and its role in human malignancies (Galeev et al., 2016, Mazumdar et al., 2015, Mullenders et al., 2015, Viny et al., 2015).

Although mice models have been created for both mitotic and meiotic specific counterparts of cohesin but homozygous deletions of mitotic counterparts is lethal as compare to heterozygous knockout mice that shows variable phenotypes. Mutations in one copy of *RAD21*, *SMC1α* or *SMC3* have been associated with cohesinopathies. The lethal phenotype of homozygous knockout mice as compared to heterozygous mice is presumed to be due to rampant aneuploidy that is incompatible with survival. Homozygous knockout of meiotic specific counterparts of cohesin *SMC1β*, *Rad21L, REC8* and *SA3* are found to be sterile while haploinsufficiency of these counterparts has not been reported.

Elucidation of the role of cohesin especially *STAG2* in myeloid malignancies and its identification as a target for therapeutic intervention needs further investigation..

Chapter Six
FUTURE WORK

Role of cohesin especially *STAG2* in myeloid malignancies raises many questions which can be addressed in future studies. Some of which are

Introduction of cohesin mutations found in MDS patients in umblical cord blood cells and studying the effect of mutations on various HSC compartments, cell cycle kinetics and cell viability.

Creating a conditional knockout murine model of *STAG2* and then determining the effect of PARP inhibitors on various haematopoietic stem cell compartments. Further conduct a gene expression profiling study on these cells to study changes in gene regulation.

Introduce *SRSF2* hotspot mutations in a myeloid cell line and study its effect on splicing of *STAG2.*

Study the interaction between CTCF and *STAG2* in knockdown cells using recent chromosome capture technologies like chromosome confirmation capture (3C)

References

Abdel-Wahab, O., Adli, M., LaFave, L. M., et al. (2012). ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer Cell*22, 180-93.

Abdel-Wahab, O., Manshouri, T., Patel, J., et al. (2010). Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res*70, 447-52.

Acquaviva, C., Gelsi-Boyer, V. & Birnbaum, D. (2010). Myelodysplastic syndromes: lost between two states? *Leukemia*24, 1-5.

Asou, H., Matsui, H., Ozaki, Y., et al. (2009). Identification of a common microdeletion cluster in 7q21.3 subband among patients with myeloid leukemia and myelodysplastic syndrome. *Biochem Biophys Res Commun*383, 245-51.

Atienza, J. M., Roth, R. B., Rosette, C., et al. (2005). Suppression of RAD21 gene expression decreases cell growth and enhances cytotoxicity of etoposide and bleomycin in human breast cancer cells. *Mol Cancer Ther*4, 361-8.

Audeh, M. W., Carmichael, J., Penson, R. T., et al. (2010). Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet*376, 245-51.

Bailey, M. L., O'Neil, N. J., van Pel, D. M., et al. (2013). Glioblastoma cells containing mutations in the cohesin component, STAG2, are sensitive to PARP inhibition. *Mol Cancer Ther*.

Bakkenist, C. J. & Kastan, M. B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature*421, 499-506.

Balbas-Martinez, C., Sagrera, A., Carrillo-de-Santa-Pau, E., et al. (2013). Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. *Nat Genet*45, 1464-9.

Bantounas, I., Phylactou, L. A. & Uney, J. B. (2004). RNA interference and the use of small interfering RNA to study gene function in mammalian systems. *Journal of Molecular Endocrinology*33, 545-557.

Barlow, J. L., Drynan, L. F., Hewett, D. R., et al. (2010). A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. *Nat Med*16, 59-66.

Bartek, J. & Lukas, J. (2007). DNA damage checkpoints: from initiation to recovery or adaptation. *Current Opinion in Cell Biology*19, 238-245.

Bejar, R., Levine, R. & Ebert, B. L. (2011a). Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J Clin Oncol*29, 504-15.

Bejar, R., Stevenson, K., Abdel-Wahab, O., et al. (2011b). Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*364, 2496-506.

Bench, A. J., Nacheva, E. P., Hood, T. L., et al. (2000). Chromosome 20 deletions in myeloid malignancies: reduction of the common deleted region, generation of a PAC/BAC contig and identification of candidate genes. UK Cancer Cytogenetics Group (UKCCG). *Oncogene*19, 3902-13.

Bennett, J. M., Catovsky, D., Daniel, M. T., et al. (1982). Proposals for the classification of the myelodysplastic syndromes. *Br J Haematol*51, 189-99.

Bermudez, V. P., Maniwa, Y., Tappin, I., et al. (2003). The alternative Ctf18-Dcc1-Ctf8-replication factor C complex required for sister chromatid cohesion loads proliferating cell nuclear antigen onto DNA. *Proc Natl Acad Sci U S A*100, 10237-42.

Biasini, M., Bienert, S., Waterhouse, A., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*42, W252-8.

Birkenbihl, R. P. & Subramani, S. (1992). Cloning and characterization of rad21 an essential gene of Schizosaccharomyces pombe involved in DNA double-strand-break repair. *Nucleic Acids Res*20, 6605-11.

Birmingham, A., Anderson, E. M., Reynolds, A., et al. (2006). 3[prime] UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Meth*3, 199-204.

Bock, F. J. & Chang, P. (2016). New Directions in PARP Biology. *FEBS J*.

Borck, G., Zarhrate, M., Bonnefont, J. P., et al. (2007). Incidence and clinical features of X-linked Cornelia de Lange syndrome due to SMC1L1 mutations. *Hum Mutat*28, 205-6.

Borges, V., Lehane, C., Lopez-Serra, L., et al. (2010). Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. *Mol Cell*39, 677-88.

Bose, T., Lee, K. K., Lu, S., et al. (2012). Cohesin proteins promote ribosomal RNA production and protein translation in yeast and human cells. *PLoS Genet*8, e1002749.

Bothmer, A., Robbiani Df Fau - Di Virgilio, M., Di Virgilio M Fau - Bunting, S. F., et al. Regulation of DNA end joining, resection, and immunoglobulin class switch recombination by 53BP1.

Boulton, S., Pemberton, L. C., Porteous, J. K., et al. (1995). Potentiation of temozolomide-induced cytotoxicity: a comparative study of the biological effects of poly(ADP-ribose) polymerase inhibitors. *British Journal of Cancer*72, 849-856.

Boultwood, J., Fidler, C., Strickson, A. J., et al. (2002). Narrowing and genomic annotation of the commonly deleted region of the 5q- syndrome. *Blood*99, 4638-41.

Boultwood, J., Perry, J., Pellagatti, A., et al. (2010). Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia*24, 1062-5.

Bowman, K. J., Newell, D. R., Calvert, A. H., et al. (2001). Differential effects of the poly (ADP-ribose) polymerase (PARP) inhibitor NU1025 on topoisomerase I and II inhibitor cytotoxicity in L1210 cells in vitro. *British Journal of Cancer*84, 106-112.

Bryant, H. E., Petermann, E., Schultz, N., et al. (2009). PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination. *EMBO J*28, 2601-15.

Bryant, H. E., Schultz, N., Thomas, H. D., et al. (2005). Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*434, 913-917.

Burgers, P. M. J., Koonin, E. V., Bruford, E., et al. (2001). Eukaryotic DNA Polymerases: Proposal for a Revised Nomenclature. *Journal of Biological Chemistry*276, 43487-43490.

Carbuccia, N., Murati, A., Trouplin, V., et al. (2009). Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia*23, 2183-6.

Carson, C. T., Schwartz, R. A., Stracker, T. H., et al. (2003). The Mre11 complex is required for ATM activation and the G2/M checkpoint. *EMBO J*22, 6610-20.

Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. (2016). Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol*26, 52-64.

Challen, G. A., Sun, D., Jeong, M., et al. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet*44, 23-31.

Chambon, P., Weill, J. D. & Mandel, P. (1963). Nicotinamide mononucleotide activation of new DNA-dependent polyadenylic acid synthesizing nuclear enzyme. *Biochem Biophys Res Commun*11, 39-43.

Chan, D. W., Chen, B. P.-C., Prithivirajsingh, S., et al. (2002). Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks. *Genes & Development*16, 2333-2338.

Chan, S. M., Thomas, D., Corces-Zimmerman, M. R., et al. (2015). Isocitrate dehydrogenase 1 and 2 mutations induce BCL-2 dependence in acute myeloid leukemia. *Nat Med*21, 178-84.

Chelysheva, L., Diallo, S., Vezon, D., et al. (2005). AtREC8 and AtSCC3 are essential to the monopolar orientation of the kinetochores during meiosis. *J Cell Sci*118, 4621-32.

Chen, C. Y., Lin, L. I., Tang, J. L., et al. (2007). RUNX1 gene mutation in primary myelodysplastic syndrome--the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. *Br J Haematol*139, 405-14.

Chen, H., Lisby, M. & Symington, L. S. (2013). RPA coordinates DNA end resection and prevents formation of DNA hairpins. *Mol Cell*50, 589-600.

Christiansen, D. H., Andersen, M. K. & Pedersen-Bjergaard, J. (2004). Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation. *Blood*104, 1474-81.

Covo, S., Westmoreland, J. W., Gordenin, D. A., et al. (2010). Cohesin Is limiting for the suppression of DNA damage-induced recombination between homologous chromosomes. *PLoS Genet*6, e1001006.

Cuadrado, A., Remeseiro, S., Grana, O., et al. (2015). The contribution of cohesin-SA1 to gene expression and chromatin architecture in two murine tissues. *Nucleic Acids Res*43, 3056-67.

Cuccato, G., Polynikis, A., Siciliano, V., et al. (2011). Modeling RNA interference in mammalian cells. *BMC Syst Biol*5, 19.

Curtin, N. (2014). PARP inhibitors for anticancer therapy. *Biochemical Society Transactions*42, 82.

Dasgupta, T., Antony, J., Braithwaite, A. W., et al. (2016). HDAC8 Inhibition Blocks SMC3 Deacetylation and Delays Cell Cycle Progression without Affecting Cohesin-dependent Transcription in MCF7 Cancer Cells. *J Biol Chem*291, 12761-70.

De Lorenzo, S. B., Patel, A. G., Hurley, R. M., et al. (2013). The Elephant and the Blind Men: Making Sense of PARP Inhibitors in Homologous Recombination Deficient Tumor Cells. *Frontiers in Oncology*3, 228.

Deardorff, M. A., Kaur, M., Yaeger, D., et al. (2007). Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of cornelia de Lange syndrome with predominant mental retardation. *Am J Hum Genet*80, 485-94.

Deardorff, M. A., Wilde, J. J., Albrecht, M., et al. (2012). RAD21 mutations cause a human cohesinopathy. *Am J Hum Genet*90, 1014-27.

Delhommeau, F., Dupont, S., Della Valle, V., et al. (2009). Mutation in TET2 in myeloid cancers. *N Engl J Med*360, 2289-301.

Diaz-Martinez, L. A., Beauchene, N. A., Furniss, K., et al. (2010). Cohesin is needed for bipolar mitosis in human cells. *Cell Cycle*9, 1764-73.

Difilippantonio, S., Celeste, A., Fernandez-Capetillo, O., et al. (2005). Role of Nbs1 in the activation of the Atm kinase revealed in humanized mouse models. *Nat Cell Biol*7, 675-85.

Ding, L., Ley, T. J., Larson, D. E., et al. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*481, 506-10.

Dorsett, D. (2009). Cohesin, gene expression and development: lessons from Drosophila. *Chromosome Res*17, 185-200.

Dorsett, D. (2011). Cohesin: genomic insights into controlling gene transcription and development. *Curr Opin Genet Dev*21, 199-206.

Dorsett, D. & Krantz, I. D. (2009). On the molecular etiology of Cornelia de Lange syndrome. *Ann N Y Acad Sci*1151, 22-37.

Dorsett, D. & Strom, L. (2012). The ancient and evolving roles of cohesin in gene expression and DNA repair. *Curr Biol*22, R240-50.

Dowdy, C. R., Xie, R., Frederick, D., et al. (2010). Definitive hematopoiesis requires Runx1 C-terminal-mediated subnuclear targeting and transactivation. *Hum Mol Genet*19, 1048-57.

Downs, J. A. & Jackson, S. P. (2004). A means to a DNA end: the many roles of Ku. *Nat Rev Mol Cell Biol*5, 367-378.

Dutt, S., Narla, A., Lin, K., et al. (2011). Haploinsufficiency for ribosomal protein genes causes selective activation of p53 in human erythroid progenitor cells. *Blood*117, 2567-76.

Ebert, B. L., Pretz, J., Bosco, J., et al. (2008). Identification of RPS14 as a 5q-syndrome gene by RNA interference screen. *Nature*451, 335-9.

El-Khamisy, S. F., Masutani, M., Suzuki, H., et al. (2003). A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic Acids Research*31, 5526-5533.

Ernst, T., Chase, A. J., Score, J., et al. (2010). Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat Genet*42, 722-6.

Esposito, M. T., Zhao, L., Fung, T. K., et al. (2015). Synthetic lethal targeting of oncogenic transcription factors in acute leukemia by PARP inhibitors. *Nat Med*21, 1481-90.

Evers, L., Perez-Mancera, P. A., Lenkiewicz, E., et al. (2014). STAG2 is a clinically relevant tumor suppressor in pancreatic ductal adenocarcinoma. *Genome Med*6, 9.

Farmer, H., McCabe, N., Lord, C. J., et al. (2005). Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*434, 917-921.

Figueroa, M. E., Abdel-Wahab, O., Lu, C., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*18, 553-67.

Fisher, J. B., Peterson, J., Reimer, M., et al. (2016). The cohesin subunit Rad21 is a negative regulator of hematopoietic self-renewal through epigenetic repression of Hoxa7 and Hoxa9. *Leukemia*.

Frizzell, K. M., Gamble, M. J., Berrocal, J. G., et al. (2009). Global Analysis of Transcriptional Regulation by Poly(ADP-ribose) Polymerase-1 and Poly(ADP-ribose) Glycohydrolase in MCF-7 Human Breast Cancer Cells. *The Journal of Biological Chemistry*284, 33926-33938.

Galeev, R., Baudet, A., Kumar, P., et al. (2016). Genome-wide RNAi Screen Identifies Cohesin Genes as Modifiers of Renewal and Differentiation in Human HSCs. *Cell Rep*14, 2988-3000.

Galm, O., Herman, J. G. & Baylin, S. B. (2006). The fundamental role of epigenetics in hematopoietic malignancies. *Blood Rev*20, 1-13.

Gaymes, T. J., Mohamedali, A. M., Patterson, M., et al. (2013). Microsatellite instability induced mutations in DNA repair genes CtIP and MRE11 confer hypersensitivity to poly (ADP-ribose) polymerase inhibitors in myeloid malignancies. *Haematologica*98, 1397-406.

Gaymes, T. J., Shall, S., Farzaneh, F., et al. (2008). Chromosomal instability syndromes are sensitive to poly ADP-ribose polymerase inhibitors. *Haematologica*93, 1886-9.

Gaymes, T. J., Shall, S., MacPherson, L. J., et al. (2009). Inhibitors of poly ADP-ribose polymerase (PARP) induce apoptosis of myeloid leukemic cells: potential for therapy of myeloid leukemia and myelodysplastic syndromes. *Haematologica*94, 638-46.

Gelmon, K. A., Tischkowitz, M., Mackay, H., et al. (2011). Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study. *Lancet Oncol*12, 852-61.

Gelot, C., Guirouilh-Barbat, J., Le Guen, T., et al. (2016). The Cohesin Complex Prevents the End Joining of Distant DNA Double-Strand Ends. *Mol Cell*61, 15-26.

Gelsi-Boyer, V., Trouplin, V., Adelaide, J., et al. (2009). Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol*145, 788-800.

Gordillo, M., Vega, H., Trainer, A. H., et al. (2008). The molecular mechanism underlying Roberts syndrome involves loss of ESCO2 acetyltransferase activity. *Hum Mol Genet*17, 2172-80.

Greenberg, P., Cox, C., LeBeau, M. M., et al. (1997). International scoring system for evaluating prognosis in myelodysplastic syndromes. *Blood*89, 2079-88.

Grisendi, S., Bernardi, R., Rossi, M., et al. (2005). Role of nucleophosmin in embryonic development and tumorigenesis. *Nature*437, 147-53.

Growney, J. D., Shigematsu, H., Li, Z., et al. (2005). Loss of Runx1 perturbs adult hematopoiesis and is associated with a myeloproliferative phenotype. *Blood*106, 494-504.

Gruber, S., Arumugam, P., Katou, Y., et al. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. *Cell*127, 523-37.

Guo, G., Sun, X., Chen, C., et al. (2013). Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet*45, 1459-63.

Gupta, P., Lavagnolli, T., Mira-Bontenbal, H., et al. (2016). Cohesin's role in pluripotency and reprogramming. *Cell Cycle*15, 324-30.

Haase, D., Germing, U., Schanz, J., et al. (2007). New insights into the prognostic impact of the karyotype in MDS and correlation with subtypes: evidence from a core dataset of 2124 patients. *Blood*110, 4385-95.

Hadjur, S., Williams, L. M., Ryan, N. K., et al. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*460, 410-3.

Haferlach, T., Nagata, Y., Grossmann, V., et al. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*28, 241-7.

Haince, J. F., McDonald, D., Rodrigue, A., et al. (2008). PARP1-dependent kinetics of recruitment of MRE11 and NBS1 proteins to multiple DNA damage sites. *J Biol Chem*283, 1197-208.

Hakimi, M. A., Bochar, D. A., Schmiesing, J. A., et al. (2002). A chromatin remodelling complex that loads cohesin onto human chromosomes. *Nature*418, 994-8.

Hanna, J. S., Kroll, E. S., Lundblad, V., et al. (2001). Saccharomyces cerevisiae CTF18 and CTF4 are required for sister chromatid cohesion. *Mol Cell Biol*21, 3144-58.

Hara, K., Zheng, G., Qu, Q., et al. (2014). Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat Struct Mol Biol*21, 864-70.

Hartwell, L. H., Szankasi, P., Roberts, C. J., et al. (1997). Integrating Genetic Approaches into the Discovery of Anticancer Drugs. *Science*278, 1064.

Hauf, S., Roitinger, E., Koch, B., et al. (2005). Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. *PLoS Biol*3, e69.

Herran, Y., Gutierrez-Caballero, C., Sanchez-Martin, M., et al. (2011). The cohesin subunit RAD21L functions in meiotic synapsis and exhibits sexual dimorphism in fertility. *EMBO J*30, 3091-105.

Herriott, A., Tudhope, S. J., Junge, G., et al. (2015). PARP1 expression, activity and ex vivo sensitivity to the PARP inhibitor, talazoparib (BMN 673), in chronic lymphocytic leukaemia. *Oncotarget*6, 43978-43991.

Horsfield, J. A., Anagnostou, S. H., Hu, J. K., et al. (2007). Cohesin-dependent regulation of Runx genes. *Development*134, 2639-49.

Hu, B., Itoh, T., Mishra, A., et al. (2011). ATP hydrolysis is required for relocating cohesin from sites occupied by its Scc2/4 loading complex. *Curr Biol*21, 12-24.

Huether, R., Dong, L., Chen, X., et al. (2014). The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes. *Nat Commun*5, 3630.

Ivanov, D. & Nasmyth, K. (2005). A topological interaction between cohesin rings and a circular minichromosome. *Cell*122, 849-60.

Ivanov, D. & Nasmyth, K. (2007). A physical assay for sister chromatid cohesion in vitro. *Mol Cell*27, 300-10.

Jackson, A. L., Bartz, S. R., Schelter, J., et al. (2003). Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotech*21, 635-637.

Jackson, A. L., Burchard, J., Schelter, J., et al. (2006). Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA*12, 1179-1187.

Jackson, A. L. & Linsley, P. S. (2004). Noise amidst the silence: off-target effects of siRNAs? *Trends Genet*20, 521-4.

Jacobs, C., Huang, Y., Masud, T., et al. (2011). A hypomorphic Artemis human disease allele causes aberrant chromosomal rearrangements and tumorigenesis. *Hum Mol Genet*20, 806-19.

Jadersten, M., Saft, L., Smith, A., et al. (2011). TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. *J Clin Oncol*29, 1971-9.

Jaju, R. J., Jones, M., Boultwood, J., et al. (2000). Combined immunophenotyping and FISH identifies the involvement of B-cells in 5q- syndrome. *Genes Chromosomes Cancer*29, 276-80.

Jazayeri, A., Falck, J., Lukas, C., et al. (2006). ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nat Cell Biol*8, 37-45.

Joslin, J. M., Fernald, A. A., Tennant, T. R., et al. (2007). Haploinsufficiency of EGR1, a candidate gene in the del(5q), leads to the development of myeloid disorders. *Blood*110, 719-26.

Kaelin, W. G. (2005). The Concept of Synthetic Lethality in the Context of Anticancer Therapy. *Nat Rev Cancer*5, 689-698.

Katainen, R., Dave, K., Pitkanen, E., et al. (2015). CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat Genet*47, 818-21.

Katis, V. L., Lipp, J. J., Imre, R., et al. (2010). Rec8 phosphorylation by casein kinase 1 and Cdc7-Dbf4 kinase regulates cohesin cleavage by separase during meiosis. *Dev Cell*18, 397-409.

Kawauchi, S., Calof, A. L., Santos, R., et al. (2009). Multiple organ system defects and transcriptional dysregulation in the Nipbl(+/-) mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet*5, e1000650.

Kenna, M. A. & Skibbens, R. V. (2003). Mechanical link between cohesion establishment and DNA replication: Ctf7p/Eco1p, a cohesion establishment factor, associates with three different replication factor C complexes. *Mol Cell Biol*23, 2999-3007.

Kilburn, L. S. & Group, T. N. T. T. M. (2008). 'Triple negative' breast cancer: a new area for phase III breast cancer clinical trials. *Clin Oncol (R Coll Radiol)*20, 35-9.

Kim, B. J., Li, Y., Zhang, J., et al. (2010). Genome-wide reinforcement of cohesin binding at pre-existing cohesin sites in response to ionizing radiation in human cells. *J Biol Chem*285, 22784-92.

Kim, E., Ilagan, J. O., Liang, Y., et al. (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell*27, 617-30.

Kim, J. S., He, X., Orr, B., et al. (2016). Intact Cohesion, Anaphase, and Chromosome Segregation in Human Cells Harboring Tumor-Derived Mutations in STAG2. *PLoS Genet*12, e1005865.

Kim, J. S., Krasieva, T. B., LaMorte, V., et al. (2002). Specific recruitment of human cohesin to laser-induced DNA damage. *J Biol Chem*277, 45149-53.

Kim, M. S., Kim, S. S., Je, E. M., et al. (2012). Mutational and expressional analyses of STAG2 gene in solid cancers. *Neoplasma*59, 524-9.

Ko, M., Huang, Y., Jankowska, A. M., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*468, 839-43.

Kon, A., Shih, L. Y., Minamino, M., et al. (2013). Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat Genet*45, 1232-7.

Kosmider, O., Gelsi-Boyer, V., Ciudad, M., et al. (2009). TET2 gene mutation is a frequent and adverse event in chronic myelomonocytic leukemia. *Haematologica*94, 1676-81.

Kosmider, O., Gelsi-Boyer, V., Slama, L., et al. (2010). Mutations of IDH1 and IDH2 genes in early and accelerated phases of myelodysplastic syndromes and MDS/myeloproliferative neoplasms. *Leukemia*24, 1094-6.

Kraus, W. L. (2008). Transcriptional Control by PARP-1: Chromatin Modulation, Enhancer-binding, Coregulation, and Insulation. *Current opinion in cell biology*20, 294-302.

Kugou, K., Fukuda, T., Yamada, S., et al. (2009). Rec8 guides canonical Spo11 distribution along yeast meiotic chromosomes. *Mol Biol Cell*20, 3064-76.

Kulasekararaj, A. G., Smith, A. E., Mian, S. A., et al. (2013). TP53 mutations in myelodysplastic syndrome are strongly correlated with aberrations of chromosome 5, and correlate with adverse prognosis. *Br J Haematol*160, 660-72.

Kumar, M. (2009). Coordinate loss of a microRNA Mir145 and a protein coding gene RPS14 co-operate in the pathogenesis of 5q- syndrome.

Kumar, R., Corbett, M. A., Van Bon, B. W., et al. (2015). Increased STAG2 dosage defines a novel cohesinopathy with intellectual disability and behavioral problems. *Hum Mol Genet*24, 7171-81.

Lai, F., Godley, L. A., Joslin, J., et al. (2001). Transcript map and comparative analysis of the 1.5-Mb commonly deleted segment of human 5q31 in malignant myeloid diseases with a del(5q). *Genomics*71, 235-45.

Langemeijer, S. M., Kuiper, R. P., Berends, M., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet*41, 838-42.

Lavin, M. F. (2008). Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer. *Nat Rev Mol Cell Biol*9, 759-69.

Le Beau, M. M., Espinosa, R., 3rd, Davis, E. M., et al. (1996). Cytogenetic and molecular delineation of a region of chromosome 7 commonly deleted in malignant myeloid diseases. *Blood*88, 1930-5.

Ledermann, J., Harter, P., Gourley, C., et al. (2012). Olaparib Maintenance Therapy in Platinum-Sensitive Relapsed Ovarian Cancer. *New England Journal of Medicine*366, 1382-1392.

Leung, R. K. M. & Whittaker, P. A. (2005). RNA interference: From gene silencing to gene-specific therapeutics. *Pharmacology & Therapeutics*107, 222-239.

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*25, 1754-60.

Li, J., Feng, W., Chen, L., et al. (2016). Downregulation of SMC1A inhibits growth and increases apoptosis and chemosensitivity of colorectal cancer cells. *J Int Med Res*44, 67-74.

Lim, Z. Y., Killick, S., Germing, U., et al. (2007). Low IPSS score and bone marrow hypocellularity in MDS patients predict hematological responses to antithymocyte globulin. *Leukemia*21, 1436-41.

Limbo, O., Chahwan, C., Yamada, Y., et al. (2007). Ctp1 is a cell-cycle-regulated protein that functions with Mre11 complex to control double-strand break repair by homologous recombination. *Mol Cell*28, 134-46.

Lin, X., Ruan, X., Anderson, M. G., et al. (2005). siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Research*33, 4527-4535.

Lindsley, R. C., Mar, B. G., Mazzola, E., et al. (2015). Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood*125, 1367-76.

Liu, L., Li, Y., Li, S., et al. (2012). Comparison of next-generation sequencing systems. *J Biomed Biotechnol*2012, 251364.

Lord, C. J., Tutt, A. N. J. & Ashworth, A. (2015). Synthetic Lethality and Cancer Therapy: Lessons Learned from the Development of PARP Inhibitors. *Annual Review of Medicine*66, 455-470.

Lorsbach, R. B., Moore, J., Mathew, S., et al. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*17, 637-41.

Losada, A. (2014). Cohesin in cancer: chromosome segregation and beyond. *Nat Rev Cancer*14, 389-93.

Losada, A. & Hirano, T. (2005). Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev*19, 1269-87.

Losman, J. A. & Kaelin, W. G., Jr. (2013). What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes Dev*27, 836-52.

Luo, H., Li, Y., Mu, J. J., et al. (2008). Regulation of intra-S phase checkpoint by ionizing radiation (IR)-dependent and IR-independent phosphorylation of SMC3. *J Biol Chem*283, 19176-83.

Luo, X. & Kraus, W. L. (2012). On PAR with PARP: cellular stress signaling through poly(ADP-ribose) and PARP-1. *Genes & Development*26, 417-432.

Makishima, H., Jankowska, A. M., Tiu, R. V., et al. (2010). Novel homo- and hemizygous mutations in EZH2 in myeloid malignancies. *Leukemia*24, 1799-804.

Malcovati, L., Della Porta, M. G., Pietra, D., et al. (2009). Molecular and clinical features of refractory anemia with ringed sideroblasts associated with marked thrombocytosis. *Blood*114, 3538-45.

Mannini, L., Cucco, F., Quarantotti, V., et al. (2013). Mutation spectrum and genotype-phenotype correlation in Cornelia de Lange syndrome. *Hum Mutat*34, 1589-96.

Mannini, L. & Musio, A. (2011). The dark side of cohesin: the carcinogenic point of view. *Mutat Res*728, 81-7.

Mayer, M. L., Gygi, S. P., Aebersold, R., et al. (2001). Identification of RFC(Ctf18p, Ctf8p, Dcc1p): an alternative RFC complex required for sister chromatid cohesion in S. cerevisiae. *Mol Cell*7, 959-70.

Mazumdar, C., Shen, Y., Xavy, S., et al. (2015). Leukemia-Associated Cohesin Mutants Dominantly Enforce Stem Cell Programs and Impair Human Hematopoietic Progenitor Differentiation. *Cell Stem Cell*17, 675-88.

McKenna, A., Hanna, M., Banks, E., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*20, 1297-303.

Mehta, G. D., Rizvi, S. M. & Ghosh, S. K. (2012). Cohesin: a guardian of genome integrity. *Biochim Biophys Acta*1823, 1324-42.

Merkenschlager, M. & Odom, D. T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. *Cell*152, 1285-97.

Mian, S. A., Rouault-Pierre, K., Smith, A. E., et al. (2015). SF3B1 mutant MDS-initiating cells may arise from the haematopoietic stem cell compartment. *Nat Commun*6, 10004.

Mian, S. A., Smith, A. E., Kulasekararaj, A. G., et al. (2013). Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. *Haematologica*98, 1058-66.

Mohamedali, A., Gaken, J., Twine, N. A., et al. (2007). Prevalence and prognostic significance of allelic imbalance by single-nucleotide polymorphism analysis in low-risk myelodysplastic syndromes. *Blood*110, 3365-73.

Moran-Crusio, K., Reavie, L., Shih, A., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*20, 11-24.

Mordes, D. A. & Cortez, D. (2008). Activation of ATR and related PIKKs. *Cell Cycle*7, 2809-12.

Moshous, D., Callebaut, I., de Chasseval, R., et al. (2003). The V(D)J recombination/DNA repair factor artemis belongs to the metallo-beta-lactamase family and constitutes a critical developmental checkpoint of the lymphoid system. *Ann N Y Acad Sci*987, 150-7.

Mufti, G. J., Bennett, J. M., Goasguen, J., et al. (2008). Diagnosis and classification of myelodysplastic syndrome: International Working Group on Morphology of myelodysplastic syndrome (IWGM-MDS) consensus proposals for the definition and enumeration of myeloblasts and ring sideroblasts. *Haematologica*93, 1712-7.

Mullenders, J., Aranda-Orgilles, B., Lhoumaud, P., et al. (2015). Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J Exp Med*212, 1833-50.

Mullighan, C. G. (2009). TET2 mutations in myelodysplasia and myeloid malignancies. *Nat Genet*41, 766-7.

Murai, J., Huang, S.-y. N., Das, B. B., et al. (2012). Differential trapping of PARP1 and PARP2 by clinical PARP inhibitors. *Cancer research*72, 5588-5599.

Murai, J., Huang, S.-y. N., Renaud, A., et al. (2014). Stereospecific PARP trapping by BMN 673 and comparison with olaparib and rucaparib. *Molecular cancer therapeutics*13, 433-443.

Musio, A., Selicorni, A., Focarelli, M. L., et al. (2006). X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nat Genet*38, 528-30.

Muto, A., Calof, A. L., Lander, A. D., et al. (2011). Multifactorial origins of heart and gut defects in nipbl-deficient zebrafish, a model of Cornelia de Lange Syndrome. *PLoS Biol*9, e1001181.

Nasmyth, K. (2011). Cohesin: a catenase with separate entry and exit gates? *Nat Cell Biol*13, 1170-7.

Neal, J. A. & Meek, K. (2011). Choosing the right path: does DNA-PK help make the decision? *Mutat Res*711, 73-86.

Nikoloski, G., Langemeijer, S. M., Kuiper, R. P., et al. (2010). Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet*42, 665-7.

Nilsson, L., Astrand-Grundstrom, I., Arvidsson, I., et al. (2000). Isolation and characterization of hematopoietic progenitor/stem cells in 5q-deleted myelodysplastic syndromes: evidence for involvement at the hematopoietic stem cell level. *Blood*96, 2012-21.

Ouyang, Z., Zheng, G., Song, J., et al. (2013). Structure of the human cohesin inhibitor Wapl. *Proc Natl Acad Sci U S A*110, 11355-60.

Paddock, M. N., Bauman, A. T., Higdon, R., et al. (2011). Competition between PARP-1 and Ku70 control the decision between high-fidelity and mutagenic DNA repair. *DNA repair*10, 338-343.

Panigrahi, A. K. & Pati, D. (2012). Higher-order orchestration of hematopoiesis: is cohesin a new player? *Exp Hematol*40, 967-73.

Panigrahi, A. K., Zhang, N., Otta, S. K., et al. (2012). A cohesin-RAD21 interactome. *Biochem J*442, 661-70.

Papaemmanuil, E., Cazzola, M., Boultwood, J., et al. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*365, 1384-95.

Papaemmanuil, E., Gerstung, M., Malcovati, L., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*122, 3616-27; quiz 3699.

Paschka, P., Schlenk, R. F., Gaidzik, V. I., et al. (2010). IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *J Clin Oncol*28, 3636-43.

Ponnaluri, V. K., Maciejewski, J. P. & Mukherji, M. (2013). A mechanistic overview of TET-mediated 5-methylcytosine oxidation. *Biochem Biophys Res Commun*436, 115-20.

Poppe, B., Dastugue, N., Vandesompele, J., et al. (2006). EVI1 is consistently expressed as principal transcript in common and rare recurrent 3q26 rearrangements. *Genes Chromosomes Cancer*45, 349-56.

Ramsden, D. A. (2011). Polymerases in nonhomologous end joining: building a bridge over broken chromosomes. *Antioxid Redox Signal*14, 2509-19.

Rathinam, C., Thien, C. B., Langdon, W. Y., et al. (2008). The E3 ubiquitin ligase c-Cbl restricts development and functions of hematopoietic stem cells. *Genes Dev*22, 992-7.

Rhodes, J. M., McEwan, M. & Horsfield, J. A. (2011). Gene regulation by cohesin in cancer: is the ring an unexpected party to proliferation? *Mol Cancer Res*9, 1587-607.

Rocquain, J., Gelsi-Boyer, V., Adelaide, J., et al. (2010). Alteration of cohesin genes in myeloid diseases. *Am J Hematol*85, 717-9.

Rogakou, E. P., Pilch, D. R., Orr, A. H., et al. (1998). DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J Biol Chem*273, 5858-68.

Rollins, R. A., Morcillo, P. & Dorsett, D. (1999). Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics*152, 577-93.

Rouleau, M., Patel, A., Hendzel, M. J., et al. (2010). PARP inhibition: PARP1 and beyond. *Nat Rev Cancer*10, 293-301.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., et al. (2008). CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A*105, 8309-14.

Saft, L., Karimi, M., Ghaderi, M., et al. (2014). p53 protein expression independently predicts outcome in patients with lower-risk myelodysplastic syndromes with del(5q). *Haematologica*99, 1041-9.

Sanada, M., Suzuki, T., Shih, L. Y., et al. (2009). Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature*460, 904-8.

Sargin, B., Choudhary, C., Crosetto, N., et al. (2007). Flt3-dependent transformation by inactivating c-Cbl mutations in AML. *Blood*110, 1004-12.

Sartori, A. A., Lukas, C., Coates, J., et al. (2007). Human CtIP promotes DNA end resection. *Nature*450, 509-14.

Schar, P., Fasi, M. & Jessberger, R. (2004). SMC1 coordinates DNA double-strand break repair pathways. *Nucleic Acids Res*32, 3921-9.

Schockel, L., Mockel, M., Mayer, B., et al. (2011). Cleavage of cohesin rings coordinates the separation of centrioles and chromatids. *Nat Cell Biol*13, 966-72.

Schubbert, S., Shannon, K. & Bollag, G. (2007). Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer*7, 295-308.

Schule, B., Oviedo, A., Johnston, K., et al. (2005). Inactivating mutations in ESCO2 cause SC phocomelia and Roberts syndrome: no phenotype-genotype correlation. *Am J Hum Genet*77, 1117-28.

Scott, B. L. & Deeg, H. J. (2010). Myelodysplastic syndromes. *Annu Rev Med*61, 345-58.

Shall, S. & de Murcia, G. (2000). Poly(ADP-ribose) polymerase-1: what have we learned from the deficient mouse model? *Mutat Res*460, 1-15.

Shen, C. H., Kim, S. H., Trousil, S., et al. (2016). Loss of cohesin complex components STAG2 or STAG3 confers resistance to BRAF inhibition in melanoma. *Nat Med*22, 1056-61.

Simbulan-Rosenthal, C. M., Ly, D. H., Rosenthal, D. S., et al. (2000). Misregulation of gene expression in primary fibroblasts lacking poly(ADP-ribose) polymerase. *Proceedings of the National Academy of Sciences of the United States of America*97, 11274-11279.

Skibbens, R. V. (2004). Chl1p, a DNA helicase-like protein in budding yeast, functions in sister-chromatid cohesion. *Genetics*166, 33-42.

Skibbens, R. V., Colquhoun, J. M., Green, M. J., et al. (2013). Cohesinopathies of a feather flock together. *PLoS Genet*9, e1004036.

Sledz, C. A., Holko, M., de Veer, M. J., et al. (2003). Activation of the interferon system by short-interfering RNAs. *Nat Cell Biol*5, 834-839.

Sloand, E. M., Wu, C. O., Greenberg, P., et al. (2008). Factors affecting response and survival in patients with myelodysplasia treated with immunosuppressive therapy. *J Clin Oncol*26, 2505-11.

Solomon, D. A., Kim, J. S., Bondaruk, J., et al. (2013). Frequent truncating mutations of STAG2 in bladder cancer. *Nat Genet*45, 1428-30.

Solomon, D. A., Kim, T., Diaz-Martinez, L. A., et al. (2011). Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*333, 1039-43.

Song, W. J., Sullivan, M. G., Legare, R. D., et al. (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet*23, 166-75.

Spitz, F., Gonzalez, F. & Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell*113, 405-17.

Starczynowski, D. T., Kuchenbauer, F., Argiropoulos, B., et al. (2010). Identification of miR-145 and miR-146a as mediators of the 5q- syndrome phenotype. *Nat Med*16, 49-58.

Steensma, D. P., Gibbons, R. J., Mesa, R. A., et al. (2005). Somatic point mutations in RUNX1/CBFA2/AML1 are common in high-risk myelodysplastic syndrome, but not in myelofibrosis with myeloid metaplasia. *Eur J Haematol*74, 47-53.

Steensma, D. P. & Tefferi, A. (2008). JAK2 V617F and ringed sideroblasts: not necessarily RARS-T. *Blood*111, 1748.

Stewart, G. S., Wang, B., Bignell, C. R., et al. (2003). MDC1 is a mediator of the mammalian DNA damage checkpoint. *Nature*421, 961-6.

Strom, L., Karlsson, C., Lindroos, H. B., et al. (2007). Postreplicative formation of cohesion is required for repair and induced by a single DNA break. *Science*317, 242-5.

Sturzenegger, A., Burdova, K., Kanagaraj, R., et al. DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells.

Sundaramoorthy, S., Vazquez-Novelle, M. D., Lekomtsev, S., et al. (2014). Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion. *EMBO J*33, 2623-42.

Symington, L. S. & Gautier, J. (2011). Double-strand break end resection and repair pathway choice. *Annu Rev Genet*45, 247-71.

Tang, J. L., Hou, H. A., Chen, C. Y., et al. (2009). AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. *Blood*114, 5352-61.

Taylor, C. F., Platt, F. M., Hurst, C. D., et al. (2013). Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. *Hum Mol Genet*.

Tefferi, A., Lasho, T. L., Abdel-Wahab, O., et al. (2010). IDH1 and IDH2 mutation studies in 1473 patients with chronic-, fibrotic- or blast-phase essential thrombocythemia, polycythemia vera or myelofibrosis. *Leukemia*24, 1302-9.

Tefferi, A. & Vardiman, J. W. (2009). Myelodysplastic syndromes. *N Engl J Med*361, 1872-85.

Thol, F., Bollin, R., Gehlhaar, M., et al. (2014). Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood*123, 914-20.

Thol, F., Kade, S., Schlarmann, C., et al. (2012). Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*119, 3578-84.

Thol, F., Weissinger, E. M., Krauter, J., et al. (2010). IDH1 mutations in patients with myelodysplastic syndromes are associated with an unfavorable prognosis. *Haematologica*95, 1668-74.

Thota, S., Viny, A. D., Makishima, H., et al. (2014). Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood*.

Tirode, F., Surdez, D., Ma, X., et al. (2014). Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov*4, 1342-53.

Tiu, R. V., Visconte, V., Traina, F., et al. (2011). Updates in cytogenetics and molecular markers in MDS. *Curr Hematol Malig Rep*6, 126-35.

Tutt, A., Robson, M., Garber, J. E., et al. (2010). Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*376, 235-44.

Unal, E., Heidinger-Pauli, J. M. & Koshland, D. (2007). DNA double-strand breaks trigger genome-wide sister-chromatid cohesion through Eco1 (Ctf7). *Science*317, 245-8.

Van den Berghe, H., Cassiman, J. J., David, G., et al. (1974). Distinct haematological disorder with deletion of long arm of no. 5 chromosome. *Nature*251, 437-8.

van der Lelij, P., Chrzanowska, K. H., Godthelp, B. C., et al. (2010). Warsaw breakage syndrome, a cohesinopathy associated with mutations in the XPD helicase family member DDX11/ChlR1. *Am J Hum Genet*86, 262-6.

Viny, A. D., Ott, C. J., Spitzer, B., et al. (2015). Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J Exp Med*212, 1819-32.

Virag, L. & Szabo, C. (2002). The therapeutic potential of poly(ADP-ribose) polymerase inhibitors. *Pharmacol Rev*54, 375-429.

Walter, M. J., Shen, D., Ding, L., et al. (2012). Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med*366, 1090-8.

Wang, J., Fernald, A. A., Anastasi, J., et al. (2010). Haploinsufficiency of Apc leads to ineffective hematopoiesis. *Blood*115, 3481-8.

Wang, L., Lawrence, M. S., Wan, Y., et al. (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*365, 2497-506.

Wang, P. W., Eisenbart, J. D., Espinosa, R., 3rd, et al. (2000). Refinement of the smallest commonly deleted segment of chromosome 20 in malignant myeloid diseases and development of a PAC-based physical and transcription map. *Genomics*67, 28-39.

Wang, Y.-Q., Wang, P.-Y., Wang, Y.-T., et al. (2016). An Update on Poly(ADP-ribose)polymerase-1 (PARP-1) Inhibitors: Opportunities and Challenges in Cancer Therapy. *Journal of Medicinal Chemistry*.

Ward, A., Hopkins, J., McKay, M., et al. (2016). Genetic Interactions Between the Meiosis-Specific Cohesin Components, STAG3, REC8, and RAD21L. *G3 (Bethesda)*6, 1713-24.

Watanabe, Y. & Nurse, P. (1999). Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature*400, 461-4.

Watrin, E. & Peters, J. M. (2009). The cohesin complex is required for the DNA damage-induced G2/M checkpoint in mammalian cells. *EMBO J*28, 2625-35.

Wei, S., Chen, X., Rocha, K., et al. (2009). A critical role for phosphatase haplodeficiency in the selective suppression of deletion 5q MDS by lenalidomide. *Proc Natl Acad Sci U S A*106, 12974-9.

Weitzer, S., Lehane, C. & Uhlmann, F. (2003). A model for ATP hydrolysis-dependent binding of cohesin to DNA. *Curr Biol*13, 1930-40.

Welch, J. S., Ley, T. J., Link, D. C., et al. (2012). The origin and evolution of mutations in acute myeloid leukemia. *Cell*150, 264-78.

Wendt, K. S., Yoshida, K., Itoh, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*451, 796-801.

Whelan, G., Kreidl, E., Wutz, G., et al. (2012). Cohesin acetyltransferase Esco2 is a cell viability factor and is required for cohesion in pericentric heterochromatin. *EMBO J*31, 71-82.

Wiktor, A., Rybicki, B. A., Piao, Z. S., et al. (2000). Clinical significance of Y chromosome loss in hematologic disease. *Genes Chromosomes Cancer*27, 11-6.

Wong, J. C., Zhang, Y., Lieuw, K. H., et al. (2010). Use of chromosome engineering to model a segmental deletion of chromosome band 7q22 found in myeloid malignancies. *Blood*115, 4524-32.

Wu, N. & Yu, H. (2012). The Smc complexes in DNA damage response. *Cell Biosci*2, 5.

Wu, Q., Ochi, T., Matak-Vinkovic, D., et al. (2011). Non-homologous end-joining partners in a helical dance: structural studies of XLF-XRCC4 interactions. *Biochem Soc Trans*39, 1387-92, suppl 2 p following 1392.

Xiao, T., Wallace, J. & Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol*31, 2174-83.

Yadav, S., Sehrawat, A., Eroglu, Z., et al. (2013). Role of SMC1 in Overcoming Drug Resistance in Triple Negative Breast Cancer. *PLoS One*8, e64338.

Yamashita, Y., Yuan, J., Suetake, I., et al. (2010). Array-based genomic resequencing of human leukemia. *Oncogene*29, 3723-31.

Yan, H., Parsons, D. W., Jin, G., et al. (2009). IDH1 and IDH2 mutations in gliomas. *N Engl J Med*360, 765-73.

Yoo, S. & Dynan, W. S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Res*27, 4679-86.

Yoshida, K., Sanada, M., Shiraishi, Y., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*478, 64-9.

Yoshida, K., Toki, T., Okuno, Y., et al. (2013). The landscape of somatic mutations in Down syndrome-related myeloid disorders. *Nat Genet*45, 1293-9.

Yun, J., Song, S. H., Kang, J. Y., et al. (2016). Reduced cohesin destabilizes high-level gene amplification by disrupting pre-replication complex bindings in human cancers with chromosomal instability. *Nucleic Acids Res*44, 558-72.

Zeng, Y., Yi, R. & Cullen, B. R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proceedings of the National Academy of Sciences of the United States of America*100, 9779-9784.

Zhang, Y. F., Jiang, R., Li, J. D., et al. (2013). SMC1A knockdown induces growth suppression of human lung adenocarcinoma cells through G1/S cell cycle phase arrest and apoptosis pathways in vitro. *Oncol Lett*5, 749-755.

Appendix

Sequence of primers for *STAG2*, *SMC1A*, *SMC3* and *RAD21*

Oligo Name	Primer Sequence
STAG2_AMP_3_F	gtagtgcgatggccagtttccgaatatttttggtgcat
STAG2_AMP_3_R	cagtgtgcagcgatgacaatgcattccccattttgtg
STAG2_AMP_4_F	gtagtgcgatggccagttgagttaaccaagcctttctttt
STAG2_AMP_4_R	cagtgtgcagcgatgaccacccccttaaaaagccatt
STAG2_AMP_5_F	gtagtgcgatggccagttgagttaaccaagcctttctttt
STAG2_AMP_5_R	cagtgtgcagcgatgaccaaatgaactgggaagaacaa
STAG2_AMP_6_F	gtagtgcgatggccagtttgattttgagaaaattagaagaagc
STAG2_AMP_6_R	cagtgtgcagcgatgactgatgcaaataatgccttagc
STAG2_AMP_7_F	gtagtgcgatggccagttgccttttattgtgtgacca
STAG2_AMP_7_R	cagtgtgcagcgatgacgcccagcctaatgcttacaa
STAG2_AMP_8_F	gtagtgcgatggccagttcatgcattctaaatgaaattgct
STAG2_AMP_8_R	cagtgtgcagcgatgactcatctcaaatctaagacaatatgcag
STAG2_AMP_9_F	gtagtgcgatggccagtgcagattagctcatttctgctta
STAG2_AMP_9_R	cagtgtgcagcgatgacccaagttggtcacacaatagc
STAG2_AMP_10_F	gtagtgcgatggccagtccccaaaatactggggaat
STAG2_AMP_10_R	cagtgtgcagcgatgactgactcagtggcactaatgga
STAG2_AMP_11_F	gtagtgcgatggccagtggcccatgcttcatttctat
STAG2_AMP_11_R	cagtgtgcagcgatgacgcagtctgaagggcacatct
STAG2_AMP_12_F	gtagtgcgatggccagtttctgaaggaatgctatggtataga
STAG2_AMP_12_R	cagtgtgcagcgatgacggaaagcaagagaaaagtgtttg
STAG2_AMP_13_F	gtagtgcgatggccagtttttttaccagtcggttcaagg
STAG2_AMP_13_R	cagtgtgcagcgatgacagctgtaaacctccatgacg
STAG2_AMP_14_F	gtagtgcgatggccagtggacgttactaaaagcacctgtt
STAG2_AMP_14_R	cagtgtgcagcgatgactgtgaaagcttcgatatgatctg
STAG2_AMP_15_F	gtagtgcgatggccagtcctacccagttctcctttactcc
STAG2_AMP_15_R	cagtgtgcagcgatgactgcctcattttaacccctttt
STAG2_AMP_16_F	gtagtgcgatggccagtttccattcttttgagttaaggcta
STAG2_AMP_16_R	cagtgtgcagcgatgactgaaaacaaaactatgcacgaa
STAG2_AMP_17_F	gtagtgcgatggccagttgggatgctgagggttttag

STAG2_AMP_17_R	cagtgtgcagcgatgacttgaaaagacattgattggcata
STAG2_AMP_18_F	gtagtgcgatggccagtcacttaacagtgctaatgggctta
STAG2_AMP_18_R	cagtgtgcagcgatgacctgaggcaactgcaacaagt
STAG2_AMP_19_F	gtagtgcgatggccagttttccctaaatgcctcacaga
STAG2_AMP_19_R	cagtgtgcagcgatgacggtttcctttcttaaaatacgttcc
STAG2_AMP_20_F	gtagtgcgatggccagtcctatcatatatgccttagttttgatg
STAG2_AMP_20_R	cagtgtgcagcgatgacactgcagtagaggggctcaa
STAG2_AMP_21_F	gtagtgcgatggccagtcccagccatattgccttaaa
STAG2_AMP_21_R	cagtgtgcagcgatgaccataacccacaaaacatgcaa
STAG2_AMP_22_F	gtagtgcgatggccagtccaaaacaattgtcattaggc
STAG2_AMP_22_R	cagtgtgcagcgatgacgcaagttgccaaaggattaca
STAG2_AMP_23_F	gtagtgcgatggccagtaaatggagacatgcctgagc
STAG2_AMP_23_R	cagtgtgcagcgatgactgtgtgagtttgctgaaaacag
STAG2_AMP_24_F	gtagtgcgatggccagtcaaaccgtctctttgttttca
STAG2_AMP_24_R	cagtgtgcagcgatgacaaggttgcttacaaaaagattacca
STAG2_AMP_25_F	gtagtgcgatggccagtggcagttagtgagaaaccttgg
STAG2_AMP_25_R	cagtgtgcagcgatgactttgattttatggtggacacaga
STAG2_AMP_26_F	gtagtgcgatggccagttgtttttccctttcaaattctc
STAG2_AMP_26_R	cagtgtgcagcgatgacggcagccatgcataaaaact
STAG2_AMP_27_F	gtagtgcgatggccagtcattctttcctgcctttgaa
STAG2_AMP_27_R	cagtgtgcagcgatgaccccaatttcaactgctacctt
STAG2_AMP_28_F	gtagtgcgatggccagttgctttctttctttccaaacag
STAG2_AMP_28_R	cagtgtgcagcgatgacttccaaatgaaagggctaga
STAG2_AMP_29_F	gtagtgcgatggccagtgcttggcaaaggaagtagtga
STAG2_AMP_29_R	cagtgtgcagcgatgactgcccttaagaatcccaaaa
STAG2_AMP_30_F	gtagtgcgatggccagtgctaacagtttcgatttcttttca
STAG2_AMP_30_R	cagtgtgcagcgatgacgaaccttaatgacaattcagttgg
STAG2_AMP_31_F	gtagtgcgatggccagttcgtcgttaattttcttttcca
STAG2_AMP_31_R	cagtgtgcagcgatgacctgcgtgcttcaaggagaat
STAG2_AMP_32_F	gtagtgcgatggccagtccaaaagattacggcctgag
STAG2_AMP_32_R	cagtgtgcagcgatgactcagacaataaggcactctcactt
STAG2_AMP_33_F	gtagtgcgatggccagtgagccacatactgctgccta

STAG2_AMP_33_R	cagtgtgcagcgatgactttgcaaaacttacttctcttgg
STAG2_AMP_34_F	gtagtgcgatggccagtcccggctgataattttccat
STAG2_AMP_34_R	cagtgtgcagcgatgaccccaggtcttcaacctcaaa
STAG2_AMP_35_F	gtagtgcgatggccagtcgagattttctcccctctctc
STAG2_AMP_35_R	cagtgtgcagcgatgacgccaactaacagcgcataaa
SMC1A_AMP_1_F	gtagtgcgatggccagttacctcagttctcgggcgta
SMC1A_AMP_1_R	cagtgtgcagcgatgacgagagtgggttcgagcagaa
SMC1A_AMP_2_F	gtagtgcgatggccagtttttgttcatctcaatatctttattc
SMC1A_AMP_2_R	cagtgtgcagcgatgacgtctagccacctcccaggac
SMC1A_AMP_3_F	gtagtgcgatggccagtgaaagctctccttgggtcct
SMC1A_AMP_3_R	cagtgtgcagcgatgacaaagcctgagatggagcaga
SMC1A_AMP_4_F	gtagtgcgatggccagtgggtaaggtgaactgggttg
SMC1A_AMP_4_F	cagtgtgcagcgatgacataaggcactggctccatcc
SMC1A_AMP_5_F	gtagtgcgatggccagtgccagtgccttatcgatgc
SMC1A_AMP_5_R	cagtgtgcagcgatgacggtttcccacaagttgcag
SMC1A_AMP_6_F	gtagtgcgatggccagtgtttcccccacccctctg
SMC1A_AMP_6_R	cagtgtgcagcgatgacatctcctccctgccaacc
SMC1A_AMP_7_F	gtagtgcgatggccagtgatccctgggctgacatct
SMC1A_AMP_7_R	cagtgtgcagcgatgacatgcttaggaccggaaacg
SMC1A_AMP_8_F	gtagtgcgatggccagttggatgagtgggcacataaa
SMC1A_AMP_8_R	cagtgtgcagcgatgactccggctggtgagaaaatac
SMC1A_AMP_9_F	gtagtgcgatggccagtgccaggagtattttctcacca
SMC1A_AMP_9_R	cagtgtgcagcgatgacgggaaaatcgtgtgacagtg
SMC1A_AMP_10_F	gtagtgcgatggccagtgattcctgagccagcttcc
SMC1A_AMP_10_R	cagtgtgcagcgatgacagtggcagaacacaaacagg
SMC1A_AMP_11_F	gtagtgcgatggccagtcaagggcccctcctctagta
SMC1A_AMP_11_R	cagtgtgcagcgatgacaagaggggagaagctgaac
SMC1A_AMP_12_F	gtagtgcgatggccagtggagctggacaggctcagta
SMC1A_AMP_12_R	cagtgtgcagcgatgacgggctaggtggtaaggtggt

SMC1A_AMP_13_F	gtagtgcgatggccagtggcagtgctgtttttagcag
SMC1A_AMP_13_R	cagtgtgcagcgatgacgagacggggaagtgaaacaa
SMC1A_AMP_14_F	gtagtgcgatggccagtgctagcccttcctgactgtg
SMC1A_AMP_14_R	cagtgtgcagcgatgacgtctgtcctgcttccagtcc
SMC1A_AMP_15_F	gtagtgcgatggccagtgtgtgtgtcagggcaggac
SMC1A_AMP_15_R	cagtgtgcagcgatgacgagctgacccttgcctatga
SMC1A_AMP_16_F	gtagtgcgatggccagtgagccctgggtctagtttcc
SMC1A_AMP_16_R	cagtgtgcagcgatgactcccatttttcctcccagtt
SMC1A_AMP_17_F	gtagtgcgatggccagttagtaggaagggtgggagca
SMC1A_AMP_17_R	cagtgtgcagcgatgacgggctaggtgagacccaaat
SMC1A_AMP_18_F	gtagtgcgatggccagtggggtgggaggacgtagata
SMC1A_AMP_18_R	cagtgtgcagcgatgacctctctcactgcccttcctg
SMC1A_AMP_19_F	gtagtgcgatggccagtgattgaaggctgcaaaggag
SMC1A_AMP_19_R	cagtgtgcagcgatgactctttccagccatggtcttc
SMC1A_AMP_20_F	gtagtgcgatggccagtaggccacactcagtcagtca
SMC1A_AMP_20_R	cagtgtgcagcgatgacgcccgtggcataccctta
SMC1A_AMP_21_F	gtagtgcgatggccagttgctctggattgtcatttgc
SMC1A_AMP_21_R	cagtgtgcagcgatgactggatggagatagggacagg
SMC1A_AMP_22_F	gtagtgcgatggccagtcctgtccctatctccatcca
SMC1A_AMP_22_R	cagtgtgcagcgatgacgattccaccaaaccatcacc
SMC1A_AMP_23_F	gtagtgcgatggccagtgctcaggcaactttgagacc
SMC1A_AMP_23_R	cagtgtgcagcgatgaccaggagttgctccctgaaac
SMC1A_AMP_24_F	gtagtgcgatggccagtcctcttctgccttctggttg
SMC1A_AMP_24_R	cagtgtgcagcgatgaccaggacctgattccctatgc
SMC1A_AMP_25_F	gtagtgcgatggccagtgctttgggcaggtatgtagg
SMC1A_AMP_25_R	cagtgtgcagcgatgacgggaaggttgggagtcaaat
SMC3_AMP_1_F	gtagtgcgatggccagtgtaggcgcctcacctgac
SMC3_AMP_1_R	cagtgtgcagcgatgacaaaggagtgccccaattctc
SMC3_AMP_2_F	gtagtgcgatggccagtaatcaccactttccaaaatgaa
SMC3_AMP_2_R	cagtgtgcagcgatgacttcctgtctccttaagccaat
SMC3_AMP_3_F	gtagtgcgatggccagtttggtttcttttgttcactgtttaat
SMC3_AMP_3_R	cagtgtgcagcgatgaccactaagctgcaaccccata

SMC3_AMP_4_F	gtagtgcgatggccagttctgcatttccttttgtcca
SMC3_AMP_4_R	cagtgtgcagcgatgacgaatctgaaacttgaacaaatgctt
SMC3_AMP_5_F	gtagtgcgatggccagtttatgccctagactttaagaatcc
SMC3_AMP_5_R	cagtgtgcagcgatgacttcccatttgttaacagagtattca
SMC3_AMP_6_F	gtagtgcgatggccagttgggcaaggactttaaaatga
SMC3_AMP_6_R	cagtgtgcagcgatgactttaggaagccgaggtggat
SMC3_AMP_7_F	gtagtgcgatggccagttgctatcaaccaaggggcta
SMC3_AMP_7_R	cagtgtgcagcgatgactctcaacatgcaatggacct
SMC3_AMP_8_F	gtagtgcgatggccagtgggatacccaaccttcatgt
SMC3_AMP_8_R	cagtgtgcagcgatgactcatggagccactctgtcac
SMC3_AMP_9_F	gtagtgcgatggccagtcataaaaatcttattcaattcttcca
SMC3_AMP_9_R	cagtgtgcagcgatgactgaaaagtgtaattatggcctga
SMC3_AMP_10_F	gtagtgcgatggccagtgaaaatttgagcagttacttttggt
SMC3_AMP_10_R	cagtgtgcagcgatgacgcccaggaagtcaaggctac
SMC3_AMP_11_F	gtagtgcgatggccagtagtgaaagagaattaatggtgtca
SMC3_AMP_11_R	cagtgtgcagcgatgacccagtcacagaaagggctgt
SMC3_AMP_12_F	gtagtgcgatggccagtcaggtgggttttctcatgaagt
SMC3_AMP_12_R	cagtgtgcagcgatgacaattgagctcctcggacaac
SMC3_AMP_13_F	gtagtgcgatggccagtttggttgcaattaaacttggtt
SMC3_AMP_13_R	cagtgtgcagcgatgacttctgaaagatttttacaaagcag
SMC3_AMP_14_F	gtagtgcgatggccagtcatccatctgcaaccactga
SMC3_AMP_14_R	cagtgtgcagcgatgacgcctgaccggttaacttgtaact
SMC3_AMP_15_F	gtagtgcgatggccagtcggtcaggcttgttttctgt
SMC3_AMP_15_R	cagtgtgcagcgatgactcacttggaacccatgacct
SMC3_AMP_16_F	gtagtgcgatggccagtgtgggctcattccttaccag
SMC3_AMP_16_R	cagtgtgcagcgatgacggaagatatgttgtacaaaaactacaa
SMC3_AMP_17_F	gtagtgcgatggccagttttgggcaacatagggagac
SMC3_AMP_17_R	cagtgtgcagcgatgaccctgggcactagcatgtgtt
SMC3_AMP_18_F	gtagtgcgatggccagttcttagttaactgtgttgtgatctctc
SMC3_AMP_18_R	cagtgtgcagcgatgaccgcccagccagattatatgt
SMC3_AMP_19_F	gtagtgcgatggccagtgggtggtttaaattactttcaattct
SMC3_AMP_19_R	cagtgtgcagcgatgactcttcccccacaatatctatca

SMC3_AMP_20_F	gtagtgcgatggccagtcatacgatatatttacaacctggaga
SMC3_AMP_20_R	cagtgtgcagcgatgacatgggtccaaacaaaaatga
SMC3_AMP_21_F	gtagtgcgatggccagtggcacctaatatttgagttgatatg
SMC3_AMP_21_R	cagtgtgcagcgatgaccgaattctttttccagtcagc
SMC3_AMP_22_F	gtagtgcgatggccagtgagcaccattcctgtgtgtg
SMC3_AMP_22_R	cagtgtgcagcgatgactaggtggcagaattccaacc
SMC3_AMP_23_F	gtagtgcgatggccagtccatagaaaatgttggcagtca
SMC3_AMP_23_R	cagtgtgcagcgatgactcaacttcctaaattccatttgc
SMC3_AMP_24a_F	gtagtgcgatggccagttcacagaaaattttaaacacaaaacc
SMC3_AMP_24a_R	cagtgtgcagcgatgactggggaagtgatccaagttc
SMC3_AMP_24b_F	gtagtgcgatggccagtcccccatcttagatttggac
SMC3_AMP_24b_R	cagtgtgcagcgatgactgagcaatttgtaacaagacttca
SMC3_AMP_25_F	gtagtgcgatggccagttgaagtcttgttacaaattgctca
SMC3_AMP_25_R	cagtgtgcagcgatgacaagagctggaagcttttgtcc
SMC3_AMP_26_F	gtagtgcgatggccagtaggatctagtctgttatccttgttct
SMC3_AMP_26_R	cagtgtgcagcgatgactgctctgaggaaaatggaat
SMC3_AMP_27_F	gtagtgcgatggccagttcttttgcaaatctgattggtg
SMC3_AMP_27_R	cagtgtgcagcgatgacacatcagcaaatattaccagctt
SMC3_AMP_28_F	gtagtgcgatggccagttcaaactcaggcagttttgaat
SMC3_AMP_28_R	cagtgtgcagcgatgacacaaagacatcagataagcaaaga
SMC3_AMP_29_F	gtagtgcgatggccagtccaagatggtgccactgc
SMC3_AMP_29_R	cagtgtgcagcgatgaccagttcctgggtatgagaatca
RAD21_AMP_2_F	gtagtgcgatggccagtttcccctcttaggttttcttctg
RAD21_AMP_2_R	cagtgtgcagcgatgaccactccaatgcccctaccta
RAD21_AMP_3_F	gtagtgcgatggccagtaaggcaactttaaaggaaaaagat
RAD21_AMP_3_R	cagtgtgcagcgatgaccaaagcatgtagatttagaaaatgtgt
RAD21_AMP_4_F	gtagtgcgatggccagttgtgcttcaaggtattttgcat
RAD21_AMP_4_R	cagtgtgcagcgatgactccaaggaaaactatacatttggaa
RAD21_AMP_5_F	gtagtgcgatggccagtggcttctaaaacttggtgggtta
RAD21_AMP_5_R	cagtgtgcagcgatgactctggttttcttttgaaattgct
RAD21_AMP_6_F	gtagtgcgatggccagtgcagttttcttttatcacagtgc
RAD21_AMP_6_R	cagtgtgcagcgatgacaaagaaattctgtttatgcgctggaa

RAD21_AMP_7_F	gtagtgcgatggccagtggtttggcatttctttttcaa
RAD21_AMP_7_R	cagtgtgcagcgatgacgaatcagaaccagtaatcataagca
RAD21_AMP_8_F	gtagtgcgatggccagtagagatttttctcctccccatt
RAD21_AMP_8_R	cagtgtgcagcgatgaccaggcaaaaaccagaaatagga
RAD21_AMP_9_F	gtagtgcgatggccagtctggcctctcaggaaaatga
RAD21_AMP_9_R	cagtgtgcagcgatgacttttagatgctcaaaatgtatatgaa
RAD21_AMP_10_F	gtagtgcgatggccagttggaagatagaaatcagtggtga
RAD21_AMP_10_R	cagtgtgcagcgatgactttctttcactctgacagtataaaggt
RAD21_AMP_11_F	gtagtgcgatggccagtctgttcaaaactgtatgacctatgtt
RAD21_AMP_11_R	cagtgtgcagcgatgacggcctccctaaataccacttg
RAD21_AMP_12_F	gtagtgcgatggccagtcttcagccactgtgtattttca
RAD21_AMP_12_R	cagtgtgcagcgatgacaagctggttggaggtttttg
RAD21_AMP_13_F	gtagtgcgatggccagtttgctcagcagcattaagtaca
RAD21_AMP_13_R	cagtgtgcagcgatgactgctttctgtttccagcatc
RAD21_AMP_14_F	gtagtgcgatggccagtctgtatcagcatcctcactgaa
RAD21_AMP_14_R	cagtgtgcagcgatgacacatgggggcaatttgtaag

Sequence of Primers used for qRT-PCR

STAG2 _F	ggggaggcgaacagtctcggggaggcgaacagtctc
STAG2_R	ttattcccaagtgaatttagtactcg
SMC1A_F	tgcactgaatgatgagattcgt
SMC1A_R	ttaattctttcatttagcaactgtctg
SMC3_F	tgcactgaatgatgagattcgt
SMC3_ R	ttaattctttcatttagcaactgtctg
STAG1_F	acatcgccctgttgctgt
STAG1_R	ctgcttgtgggtcatgtctg
GAPDH_F	agccacatcgctcagaca
GAPDH_R	gcccaatacgaccaaatcc
β-Tubulin_F	ggaggagctggcaacaact
β-Tubulin_R	atctgcctcccggtctatg

Sequence for *STAG2* shRNAs

shRNA 1	ccactgatgtcttaccgaaatctcgagatttcggtaagacatcagtggt
shRNA 2	gcagttcttacagctttgtttctcgagaaacaaagctgtaagaactgct